

THE COSMOLOGICAL IMPACT OF LUMINOUS TEV BLAZARS III:
IMPLICATIONS FOR GALAXY CLUSTERS AND THE FORMATION OF DWARF GALAXIESCHRISTOPH PFROMMER^{1,2}, PHILIP CHANG^{2,3}, AND AVERY E. BRODERICK^{2,4,5}¹ Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg, Germany; christoph.pfrommer@h-its.org
² Canadian Institute for Theoretical Astrophysics, 60 St. George Street, Toronto, ON M5S 3H8, Canada; aeb@cita.utoronto.ca, pchang@cita.utoronto.ca³ Department of Physics, University of Wisconsin-Milwaukee, 1900 E. Kenwood Boulevard, Milwaukee, WI 53211, USA⁴ Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON, N2L 2Y5, Canada⁵ Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada*Draft version May 10, 2012*

ABSTRACT

A subset of blazars are powerful TeV emitters, dominating the extragalactic component of the very high energy gamma-ray universe ($E \gtrsim 100$ GeV). These TeV gamma rays generate ultra-relativistic electron-positron pairs via pair production with the extragalactic background light. While it has generally been assumed that the kinetic energy of these pairs cascade to GeV gamma rays via inverse Compton scattering, we have argued in Broderick et al. (2012, Paper I in this series) that plasma beam instabilities are capable of dissipating the pairs' energy locally on timescales short in comparison to the inverse-Compton cooling time, heating the intergalactic medium (IGM) with a rate that is independent of density. This dramatically increases the entropy of the IGM after redshift $z \sim 2$, with a number of important implications for structure formation: (1) this suggests a scenario for the origin of the cool core (CC)/non-cool core (NCC) bimodality in galaxy clusters and groups. Early-forming galaxy groups are unaffected because they can efficiently radiate the additional entropy, developing a CC. However, late-forming groups do not have sufficient time to cool before the entropy is gravitationally reprocessed through successive mergers—counteracting cooling and potentially raising the core entropy further. This may result in a population of X-ray dim groups/clusters, consistent with X-ray stacking analyses of optically selected samples. Hence blazar heating works different than feedback by active galactic nuclei, which we show can balance radiative cooling but is unable to transform CC into NCC clusters on the buoyancy timescale due to the weak coupling between the mechanical energy to the cluster gas. (2) We predict a suppression of the Sunyaev-Zel'dovich (SZ) power spectrum template on angular scales smaller than 5' due to the globally reduced central pressure of groups and clusters forming after $z \sim 1$. This allows for a larger rms amplitude of the density power spectrum, σ_8 , and may reconcile SZ-inferred values with those by other cosmological probes even after allowing for a contribution due to patchy reionization. (3) Our redshift dependent entropy floor increases the characteristic halo mass below which dwarf galaxies cannot form by a factor of approximately 10 (50) at mean density (in voids) over that found in models that include photoionization alone. This prevents the formation of late-forming dwarf galaxies ($z \lesssim 2$) with masses ranging from 10^{10} to $10^{11} M_{\odot}$ for redshifts $z \sim 2$ to 0, respectively. This may help resolve the “missing satellites problem” in the Milky Way of the low observed abundances of dwarf satellites compared to cold dark matter simulations and may bring the observed early star formation histories into agreement with galaxy formation models. At the same time, it explains the “void phenomenon” by suppressing the formation of galaxies within existing dwarf halos of masses $< 3 \times 10^{10} M_{\odot}$ with a maximum circular velocity < 60 km s⁻¹ for $z \lesssim 2$; hence reconciling the number of dwarfs in low-density regions in simulations and the paucity of those in observations.

Subject headings: BL Lacertae objects: general – galaxies: clusters: general – galaxies: formation – galaxies: dwarf – gamma rays: general – intergalactic medium

1. INTRODUCTION

Extragalactic relativistic jets are powered by accreting supermassive black holes (or in general the engines of active galactic nuclei, AGNs) and are able to carry an enormous amount of power out to cosmological distances. Blazars are a subclass of AGNs where the jet opening angle of typically $\sim 10^{\circ}$ contains our line-of-sight, allowing us to detect the Doppler-boosted radiation. Blazars are the dominant extragalactic source class in the TeV sky with currently 36 known objects out of 46 extragalactic sources (of the remaining 10, 4 are radio galaxies, 2 are starburst galaxies, and 4 are not yet identified; for a review, see Hinton & Hofmann 2009).¹ Recent observations by the *Fermi* Space Telescope and ground based imaging atmo-

spheric Cherenkov telescopes (H.E.S.S., MAGIC, and VERITAS)² demonstrated that most of the electromagnetic power is emitted in the gamma-ray band. The most extreme blazars achieve energies of up to 10 TeV, giving rise to the class of high-energy peaked BL Lac objects (HBL) while the somewhat less efficient accelerators in intermediate-energy peaked BL Lac objects (IBL) are, in some cases, also able to reach energies beyond 100 GeV. The emission mechanism is thought to be inverse Compton scattering of ultra-relativistic electrons in the jet giving rise to power-law energy spectra that increase as a function of energy and peak at the maximum energy that the accelerating process of the radiating relativistic electrons is able to deliver. The universe is not transparent to very-high energy

¹ For an up-to-date list/visualization of the extragalactic TeV sky, see <http://www.mppmu.mpg.de/~rwagner/sources/> or <http://tevcat.uchicago.edu/>.

² High Energy Stereoscopic System, Major Atmospheric Gamma Imaging Cerenkov Telescope, and Very Energetic Radiation Imaging Telescope Array System.

gamma-ray radiation (VHEGR; $E \gtrsim 100$ GeV), i.e., a beam of these energetic photons will necessarily produce electron and positron pairs off of the extragalactic background light (EBL), with typical mean free paths of VHEGRs ranging from 30 Mpc to 1 Gpc depending upon gamma-ray energy and source redshift. The pairs produced by VHEGR radiation have typical Lorentz factors of $10^5 - 10^7$.

There are only two possible ultimate destinations for the kinetic energy of the pairs: *the energy can either be channeled into lower energy (GeV) gamma-ray radiation (to which the universe is transparent), or heat the ambient medium with a partitioning factor that depends on the relative rates of the processes.* The first process has generally been assumed to dominate the manner in which these pairs lose energy, almost exclusively through inverse Compton scattering the cosmic microwave background (CMB) and EBL on a typical mean free path of $(10 - 100)$ kpc today. When the up-scattered gamma-ray is itself a VHEGR the process repeats, creating a second generation of pairs and up-scattering additional photons. The result is an inverse Compton cascade depositing the energy of the original VHEGR in gamma rays with energies $\lesssim 100$ GeV.

There are, however, problems with this picture. First, the expected inverse Compton bump has not been seen in the spectra of luminous blazars around 10 GeV. This could imply the existence of intergalactic magnetic fields that deflect the pairs out of our line-of-sight, hence reducing the inverse Compton emission (Neronov & Vovk 2010; Tavecchio et al. 2010, 2011; Dermer et al. 2011; Taylor et al. 2011; Dolag et al. 2011; Takahashi et al. 2012; Vovk et al. 2012). Typical lower limits for magnetic fields would then range from 10^{-19} G to 10^{-15} G, depending on the assumed duty cycle of blazars and are dominated by void regions, which dominate a typical line-of-sight. The values at the upper end are of astrophysical interest in the context of the formation of galactic fields³. Second, the spectral shape of the unresolved extragalactic gamma-ray background (EGRB), that has been measured by *Fermi*, exhibits a steep power-law at energies below 100 GeV (Abdo et al. 2010). If blazars contribute substantially to the EGRB, we would expect to see a flattening toward high energies, due to the inverse Compton cascades, in conflict with the *Fermi* EGRB. Traditionally, the EGRB is then used to constrain the evolution of the luminosity density of VHEGR sources (see, e.g., Narumoto & Totani 2006; Kneiske & Mannheim 2008; Inoue & Totani 2009; Venters 2010). Generally, it has been found that these cannot have exhibited the dramatic rise in numbers by $z \sim 1 - 2$ seen in the quasar distribution. That is, the co-moving number of blazars must have remained essentially fixed, at odds with both the large-scale mass assembly history of the universe, e.g., star formation history, and with the luminosity history of similarly accreting systems, e.g., the quasar luminosity density.

In our first companion paper of this series (Broderick et al. 2012, hereafter Paper I), we argued that instead of initiating an inverse Compton cascade, the pairs dissipate their kinetic energy locally, heating the intergalactic medium (IGM). We identified a process that operates on a timescale fast in comparison to inverse Compton cooling, dominating the latter for luminous TeV blazars, i.e., for HBL and IBL blazars with an equivalent isotropic luminosity of $L \gtrsim 10^{42}$ erg s⁻¹ above 100 GeV. Despite its dilute nature, the VHEGR-generated beam of ultra-

relativistic pairs, propagating through the IGM, is susceptible to plasma beam instabilities. While the commonly discussed Weibel and two-stream instabilities are strongly suppressed by finite beam temperatures, these are special cases of a general filamentary “oblique” instability which is far more virulent, and strongly insensitive to finite temperature effects (Bret et al. 2004, 2005; Bret 2009; Bret et al. 2010; Lemoine & Pelletier 2010).

If these instabilities saturate at a rate that is comparable to their linear growth rate, the beam kinetic energy is directly transferred to electrostatic modes which rapidly dissipate locally, heating the IGM. If this scenario or an analogous, similarly efficient mechanism operates in practice, it necessarily suppresses the inverse Compton cascades and naturally explains the absence of an inverse Compton bump in the TeV blazar spectra without invoking an intergalactic magnetic field. At the same time, it allows for a redshift evolution of TeV blazars that is identical to that of quasars without overproducing the EGRB. In fact, for plausible parameters of TeV blazar spectra, it is possible to explain the high-energy part of the EGRB (Paper I) without the need to appeal to exotic phenomena (e.g., dark matter annihilation, Cavadini et al. 2011).

By dissipating the pairs’ energy into the IGM, plasma instabilities (or similar processes) provide a novel mechanism for heating the universe. Integrating over the energy flux per mean free path of all known TeV blazars yields a luminosity density, or equivalently a local heating rate, that dominates that of photoheating by more than an order of magnitude at the present epoch, after accounting for incompleteness corrections (Chang et al. 2012, hereafter Paper II). We have demonstrated that the local TeV blazar luminosity function is consistent with a scaled version of the quasar luminosity function (Hopkins et al. 2007), thus the conservative assumption is that they evolve similarly, presumably due to the same underlying accretion physics. With this assumption, we showed in Paper II that for the redshifts at which blazar heating is likely to be important, $z \lesssim 3.5$, the heating rate will be relatively uniform throughout space. Between $z \sim 3.5$ and 6 it may experience order 50% fluctuations, and by $z \gtrsim 6$, it will exhibit significant stochasticity with order unity deviations.

This heating differs from other feedback prescriptions in an important way: since the number density of EBL photons and that of TeV blazars are nearly homogeneously distributed on cosmological scales, so is the resulting pair density. Hence, the implied heating rate is also homogeneous, i.e., the volumetric blazar heating rate is uniform and independent of IGM density⁴. Due to the large mean free path of TeV-photons, which is much larger than the turn-around radius of any virialized structures, the heating process is not expected to be dominated by contributions in highly biased regions at late times for $z \lesssim 3.5$. (At early times, when blazar heating exhibits considerable fluctuations, clustering bias could substantially modify the phenomenology of the heating mechanism.) The effect of a uniform heating rate is that the energy deposited per baryon is substantially larger in more tenuous regions of the universe. As a result, underdense regions experience a larger temperature increase, producing an inverted temperature-density relation in voids, asymptotically approaching $T \propto \rho_g^{-1}$. Generally, we found in Paper II that without any fine tuning it is possible to reproduce the inverted temperature-density relation at

³ After adiabatic contraction and a handful of windings, nG field strengths can be produced from an intergalactic magnetic field of $\sim 10^{-15}$ G.

⁴ The term “blazar heating” exclusively denotes the dissipation of gamma-ray luminosity of HBL and IBL blazars between 100 GeV and 10 TeV with an equivalent isotropic luminosity of $L \gtrsim 10^{42}$ erg s⁻¹ (see Paper II for details).

$z = 2 - 3$ inferred by high-redshift Ly α studies (Bolton et al. 2008; Viel et al. 2009), while simultaneously satisfying the temperature constraints at $z = 2$ (e.g., those by Lidz et al. 2010) and leaving the local Ly α forest unaffected.

In a follow-up paper by Puchwein et al. (2011), we used hydrodynamic simulations of cosmological structure formation to explicitly demonstrate that blazar heating provides not only an excellent description of the one- and two-point statistics of high-redshift Ly α forest spectra but also the line width and column density distribution. This detailed agreement includes reproducing the observed mean transmission, and is achieved using the most recent estimate of the evolution of the photoionizing background without any tuning. These successes are due specifically to the salient properties of blazar heating, in particular its excess energy injection into the low density IGM and its continuous nature.

In this work, we are interested in the impact of such an important heating mechanism on cosmological structure formation. While we propose a well-motivated, physical heating mechanism that uses a certain type of plasma instability, our conclusions concerning the thermal history of the IGM, the Ly α forest, as well as structure formation remain robust in that they only rely on a mechanism that dissipates the energy of TeV blazars independently of the density and employs a redshift evolution similar to that of the quasar luminosity function. To study the impact on structure formation, we turn to the evolution of the entropy because in the absence of radiative cooling, entropy is conserved upon adiabatic compression and can only be increased through dissipation of gravitational energy in structure formation shocks or heating due to photoionization. Hence, entropy is a unique thermodynamic variable with which to learn about the impact of non-gravitational feedback processes. We identify two different classes of objects where the time variable minimum entropy, or entropy floor, of the IGM induced by blazar heating might dramatically change our present picture of structure formation: the structure of galaxy groups and clusters, and the formation of dwarf galaxies. Before addressing the consequences of blazar heating for each, we introduce each separately, highlighting the relevant outstanding problems.

1.1. The entropy problem in galaxy groups

In the absence of non-gravitational energy injection, the X-ray luminosity of clusters (L_x) is expected to exhibit a self-similar scaling with intracluster medium (ICM) temperature (T) as $L_x \propto T^2$ from purely gravitational and shock dynamics (Kaiser 1986; Evrard & Henry 1991; Evrard et al. 1996). However, this cannot extend from clusters down to groups without vastly overproducing the extragalactic soft X-ray background. Indeed, groups are observed to have a smaller X-ray luminosity compared to the self-similar expectation yielding a steeper scaling with temperature, $L_x \propto T^3$ (Markevitch 1998). This can be explained if the gas was initially preheated to an entropy floor of $\sim 100 \text{ keV cm}^2$ which reconciles the simulated background with observations (see Voit 2005, for a review).

In principle, there are three physical processes that could produce such an entropy floor for groups, all of which may occur in practice, raising the question of which, if any, is dominant: (1) catastrophic cooling and collapse of the low-entropy gas at the centers of halos, allowing accretion-shock heated gas to adiabatically flow inward and replace the condensed gas with its elevated entropy level (Voit & Bryan 2001a,b; Voit et al. 2003), (2) an early epoch of global en-

tropy injection prior to the formation of groups and clusters, typically referred to as “preheating” (e.g., Kaiser 1991; Evrard & Henry 1991; Ponman et al. 1999; Balogh et al. 1999; Pen 1999; Borgani et al. 2001; Bryan & Voit 2001; Croft et al. 2001; Bialek et al. 2001; Babul et al. 2002; Voit et al. 2003, 2005; Borgani & Viel 2009; Stanek et al. 2010), and (3) self-regulated AGN feedback at the cores of groups and clusters (e.g., Churazov et al. 2001; Sijacki & Springel 2006; Sijacki et al. 2007, 2008; McNamara & Nulsen 2007; Puchwein et al. 2008; Booth & Schaye 2009; McCarthy et al. 2010; Dubois et al. 2010; Teyssier et al. 2011).

Catastrophic cooling has been found to be unstable in large-scale numerical hydrodynamic simulations, typically resulting in an untenably large fraction of baryons in stars (see Borgani & Kravtsov 2009, for a review). More importantly, runaway cooling generally precludes the existence of the observed hot groups, making it clear that it cannot be the sole explanation of the low X-ray luminosity of groups. However, radiative cooling is a physical process that will occur if it is not perfectly balanced by some heating process and the observations of multiphase gas in cool cores (CCs) suggest that cooling is happening (Donahue et al. 2011), presumably through thermal instability (McCourt et al. 2012). Hence, while *some* of the elevated entropies in groups have to be attributed to cooling, the absence of a cooling catastrophe implies a (non-gravitational) energy feedback process that will inevitably also raise the core entropy.

In contrast, preheating has met with broad success in explaining the observed $L_x - T$ correlation, ostensibly by suppressing the formation of dense cluster cores at low masses and therefore reducing L_x for group-sized objects. Typically, the redshift range adopted for the injection of the entropy, $z \gtrsim 3$, ensures that preheating is complete well before any galaxy group or cluster has turned around from the Hubble expansion and started to collapse. Hence the entropy, $K_e = kTn_e^{-2/3}$ (where n_e is the electron density and k is the Boltzmann constant), is generally injected at the lowest possible gas density (and thus temperature) at that redshift. This minimizes the heating necessary to produce the observed impact upon the entropy, typically 1 keV per particle. Numerical simulations incorporating a minimum entropy of the gas, or entropy floor, of $K_e \simeq (100 - 200) \text{ keV cm}^2$ at redshifts around $z \simeq 3 - 4$, are able to steepen the $L - T$ relation from the purely gravitational collapse estimate ($L_x \propto T^2$), and find broad agreement with observations (Kaiser 1991; Balogh et al. 1999; Bialek et al. 2001; Voit et al. 2002, 2003; Stanek et al. 2010). Nevertheless, there are serious observational difficulties confronting the standard preheating scenarios, apart from the theoretical problem of lacking a physical model for the heating process. First, after preheating, star formation in L_* and lower mass galaxies is significantly suppressed (e.g., Oh & Benson 2003; Benson & Madau 2003) and hence inconsistent with observational data that finds, e.g., the peak of the star formation rate at redshifts $z \sim 2$. Second, the presence of a high entropy floor everywhere in the universe makes it impossible for groups to radiate away the excess entropy and impossible to explain the existence of a subset of X-ray luminous, CC groups with steep entropy profiles, and low values of the central entropy (McCarthy et al. 2008; Fang & Haiman 2008; Cavagnolo et al. 2009).

Heating via feedback processes, e.g., from star formation (stellar winds and supernova) and AGNs (for a review, see McNamara & Nulsen 2007), has already been shown to have a large impact on the formation and history of galaxy clus-

ters and groups. Self-regulated, inhomogeneous energy feedback mechanisms are very successful in globally stabilizing the group and cluster atmospheres, and in particular, preventing the cooling catastrophe. The resulting gas mass profiles, gas fractions, and the $L_x - T$ correlation compare impressively to those observed (Sijacki et al. 2008; Puchwein et al. 2008). While there seems to be a globally convergent scenario emerging with convincing energetics and duty cycles, the actual heating process has yet to be identified and may involve interesting astrophysics, e.g., turbulence (Enßlin & Vogt 2006), cosmic rays (Guo & Oh 2008; Enßlin et al. 2011), or plasma instabilities (Kunz et al. 2011).

As we will show in this paper, the entropy injection by blazars is in some sense an amalgam of both the preheating and feedback mechanisms. Unlike typical AGN feedback, the effect of the blazar heating is not localized, operating at much larger distances typically and thus over considerably longer timescales. Unlike instantaneous preheating, the blazars provide a *time-dependent* entropy injection rate, peaking near $z \sim 1$. As a consequence, we will show that the formation of L_* galaxies is not suppressed and early-forming groups are little affected by blazar heating, having had time to cool and develop an X-ray luminous CC; avoiding the primary difficulties with the standard preheating scenario. On the other hand, the blazar-heated ICM ending up in late-forming groups will not have had sufficient time to radiate the additional entropy away before it can be gravitationally reprocessed in merging shocks which are ubiquitous in a hierarchically growing universe. We will argue that this can lead to elevated entropy core values resembling those of *non-cool core (NCC) clusters*.

1.2. The dwarf problem in our Galaxy and nearby voids

The Λ cold dark matter (Λ CDM) concordance cosmology predicts that Milky Way-sized halos should contain many more dwarf-sized halos than the observed number of dwarf galaxies, the so-called “substructure problem”, or within the local context the “missing satellites problem” (for a recent review, see Kravtsov 2010). Closely related to this problem is the “void phenomenon” which is the apparent discrepancy between the number of dwarfs in low-density regions in simulations and the paucity of those in observations (Peebles 2001). In principle, both problems can be solved in three general ways: by suppressing the formation of dwarf halos, suppressing the formation of galaxies within existing dwarf halos, and/or suppressing star formation within dwarf galaxies. In the case of the “void phenomenon”, there is a forth class of models that solves the problem. Considering interacting dark matter and dark energy as mediated, e.g., by a Yukawa coupling, implies a fifth force that reduces late accretion onto halos and pushes matter out of voids, hence resolving the discrepancy of dwarf abundances in voids (Farrar & Peebles 2004; Nusser et al. 2005). In the following, we review the three classes of models that are able to solve both problems simultaneously.

Suppressing the formation of dwarf halos is difficult to accomplish within the context of Λ CDM, requiring modifications to the standard cosmological paradigm, such as interacting dark matter (Spergel & Steinhardt 2000), modifications to the seed perturbation spectrum (Zentner & Bullock 2003), or warm dark matter (WDM Dalcanton & Hogan 2001; Macciò & Fontanot 2010). While we will discuss the last of these in Section 3.3, given the current success the Λ CDM model has had predicting the observed halo structures (see, e.g., Dalal & Kochanek 2002; Mao et al. 2004), we will not comment upon these possibilities further.

Suppressing the formation of dwarf galaxies, i.e., preventing the accretion of baryons by existing dwarf halos, may be accomplished in principle by a variety of mechanisms, including photoionization heating (Efstathiou 1992; Kauffmann et al. 1993; Quinn et al. 1996; Thoul & Weinberg 1996; Kitayama & Ikeuchi 2000; Bullock et al. 2000, 2001; Chiu et al. 2001; Somerville 2002; Dijkstra et al. 2004) or accretion shock heating (Scannapieco et al. 2001; Kravtsov et al. 2004; Sigward et al. 2005). If baryons can collect, they may not be able to cool efficiently due to a lack of H I as a result of photoionization (Haiman et al. 1997, 2000) or intrinsically low metallicities (Kravtsov et al. 2004; Kaufmann et al. 2007; Tassis et al. 2008; Robertson & Kravtsov 2008; Gnedin et al. 2009). Nevertheless, recent numerical simulations which self-consistently include the photoionizing background due to star formation have found that while it does have a pronounced effect, it cannot suppress dwarf galaxy formation at the level implied by observations (Hoefl et al. 2006; Okamoto et al. 2008; Nickerson et al. 2011). More importantly, the metallicity distribution of some dwarfs is inconsistent with dwarfs generally being pre-reionization relics (Dolphin et al. 2005; Fenner et al. 2006; Holtzman et al. 2006; Orban et al. 2008), and therefore the suppression of dwarf formation must have occurred more recently than the epoch of reionization.

Once formed, the gas may be removed from dwarf galaxies via photo-evaporation (Barkana & Loeb 1999; Shapiro et al. 2004) and feedback from the first supernovae (Mac Low & Ferrara 1999; Dekel & Woo 2003; Mashchenko et al. 2008; Jubelgas et al. 2008; Wadehuh & Springel 2011; Nickerson et al. 2011; Uhlig et al. 2012). Tidal interactions of satellite dwarf halos with the Milky Way may result in a dramatic decrease in their mass, and to a lesser extent in circular velocity, after $z \sim 2$ (Kravtsov et al. 2004; Nickerson et al. 2011). However, these processes also strip material from larger halos, with the result that the smallest dwarf spheroidal galaxies presently within the Local Group may have had a mass at formation that was much larger than currently observed, and thus were capable of building up a sizeable stellar component in their seemingly shallow potential wells.

As we will show in this work, the heating due to blazars provides an additional mechanism to suppress the formation of dwarfs. Unlike photoionization models, which typically invoke the heating at reionization, blazar heating provides a well defined, time-dependent suppression mechanism, with the suppression rising dramatically after $z \sim 2$. In addition, due to its insensitivity to density, the heating from blazars suppresses structure formation most efficiently in the low-density regions that are responsible for late-forming dwarf halos. As a result, the impact from blazars is not degenerate with variations in the parameters of reionization and/or tidal interactions.

1.3. Structure of this Paper

This is the third in a series of three papers that discuss the potential cosmological impact of TeV emission from blazars. Paper I provides a plausible mechanism for the local dissipation of the TeV luminosity, effectively producing an additional heating process within the IGM, and its implications upon high-energy gamma-ray observations. Paper II estimates the magnitude of the new heating term, describes the associated modifications to the thermal history of the IGM, and shows how this can explain some recent observations of the Ly α forest. Paper III, this paper, considers the impact the new heating term will have upon the structure and statistics of galaxy clusters

and groups, and upon the ages and properties of dwarf galaxies throughout the universe, generally finding that blazar heating can help explain outstanding questions in both cases.

In Section 2, we show that blazar heating necessarily implies the injection of a tremendous amount of entropy and employ the implication for the evolution of the warm and hot phases of the IGM. In particular, we show the broad implications of this heating on the X-ray population of clusters (Section 2.1), the entropy structure of galaxy groups/clusters (Section 2.2), the CC/NCC cluster bimodality (Section 2.3), and its impact on the Sunyaev-Zel'dovich (SZ) power-spectrum (Section 2.4). We discuss the impact of this heating on the formation of structure in the universe in Section 3, showing from the viewpoint of linear theory that the formation of dwarf galaxies that end up in Milky Way-sized halos (Section 3.1) as well as those in voids (Section 3.2) are suppressed at late redshift ($z \lesssim 2$). Finally, we conclude in Section 4.

The calculations presented below assume the WMAP7 cosmology (Komatsu et al. 2011) with $h_0 = 0.704$, $\Omega_{\text{DM}} = 0.227$, $\Omega_B = 0.0456$, $\Omega_{\Lambda} = 0.728$, $\sigma_8 = 0.81$, $n_s = 0.967$, and a matter transfer function that accounts for the baryonic features (Eisenstein & Hu 1998).

2. THE EVOLUTION OF ENTROPY AND THE HOT PHASE OF THE INTERGALACTIC MEDIUM

The injection of large amounts of energy into the IGM by blazars is accompanied by a substantial increase in the IGM entropy. Here, we discuss the impact of this additional entropy upon the evolution of the gravitationally heated warm and hot components of the IGM. In doing so we neglect radiative cooling, an approximation that is well justified in the low-density regions of the IGM generally and one that we will justify for late-forming groups and clusters below. In this case the entropy of the IGM necessarily increases during gravitational collapse (e.g., due to structure-formation shocks, feedback, and/or photoheating). As a consequence, the entropy injected by TeV blazars places an elevated floor upon the entropy of late-forming structures within the IGM. If radiative cooling is permitted (e.g., at the centers of early-forming clusters), the gas entropy may decrease,⁵ reducing the impact of the blazar heating. In this work, we will encounter two different definitions of entropy⁶ which are proportional to each other, namely,

$$K_e \equiv \frac{kT}{n_e^{2/3}} \quad \text{and} \quad K \equiv \frac{P}{\rho_g^{5/3}} = \frac{kT}{\mu m_p \rho_g^{2/3}}. \quad (1)$$

K_e is a quantity conveniently used in the X-ray literature as it can be directly constructed from the observables temperature, T , and electron density, n_e . The quantity K is the constant of proportionality in the equation of state $P(\rho_g)$ for an adiabatic monatomic gas with mass density ρ_g and pressure P .

The evolution of the temperature and the entropy are shown in Figure 1 for a fluid element at overdensity $\delta = \rho/\bar{\rho} - 1 = 0$ (where $\bar{\rho}$ and ρ denote the average and local matter density) for the cases of pure photoheating and two realizations of blazar

⁵ According to the second law of thermodynamics, the entropy of an isolated system which is not in equilibrium has to increase over time and can only remain constant during a reversible process. In the case of radiative cooling, the gas shares its entropy with the cosmic radiation field—a process that effectively removes entropy from the gaseous phase. However, the total entropy of the system including gas and radiation increases in this process.

⁶ These definitions of entropy are related to the standard thermodynamic definition of entropy per particle by $s = k \ln K^{3/2} + s_0$, where s_0 is a constant that depends upon fundamental constants and a mixture of particle masses.

heating. Our standard model is normalized to the observed number of TeV blazars, accounts for an incompleteness correction for incomplete sky coverage at TeV energies and duty cycle, employs conservative assumptions of spectral corrections and contributing source classes for blazar heating, and has a heating rate of $7 \times 10^{-8} \text{ eV cm}^{-3} \text{ Gyr}^{-1}$ at $z = 0$ (Paper II). The heating rate in the optimistic model is a factor two larger, implying a rate of $1.4 \times 10^{-7} \text{ eV cm}^{-3} \text{ Gyr}^{-1}$ at $z = 0$ and matches the inverted temperature-density relation found in high-redshift Ly α observations (Bolton et al. 2008; Viel et al. 2009; Puchwein et al. 2011). Since any mechanism that injects energy will produce a corresponding increase in K_e , in the absence of cooling the entropy accumulates monotonically regardless of the mechanism responsible for heating the IGM. Nevertheless, some general statements can be made about the distinction between photoheating and that due to blazars.

Due to the ionization balance maintained between recombination and photoionization, photoheating produces a generic entropy injection profile following the epoch of reionization; this is simply the “loss of memory” effect (Hui & Gnedin 1997; Hui & Haiman 2003). In the absence of the blazar component, photoheating of hydrogen is the main source of heating for the universe at mean density at late times. The heating rate is given by the rate at which hydrogen recombines and is then reionized. In photoionization equilibrium, the temperature of the system will adjust such that the timescales of the net photoheating and recombination cooling rates become equal implying a steady state. Since recombination does not depend on the UV photon density the rate of photoionization is also independent of the UV photon density. Hence the late time photoheating of the IGM is independent of the number of ionizing photons, and thus the number of ionizing sources (above the ionization threshold). As a consequence, the photoheating contribution to the IGM entropy at $z = 0$ is limited to near $K_e \approx 8 \text{ keV cm}^2$ for $\delta = 0$.

Such an argument does not work for blazar heating. Here the efficiency is close to the maximum of 100% all the time⁷ provided there are sufficient numbers of EBL photons and each point in space is reached by a number of blazar beams of TeV photons (which is likely the case; see Section 3.2 of Paper II). Therefore, the heating rate, and thus temperature of the IGM, is nearly linearly dependent upon the VHEGR luminosity density of TeV blazars. Correspondingly, the entropy injection from blazars depends sensitively upon the history of the TeV blazar population, and thus the cumulative contribution of blazars to the present-day K_e is somewhat uncertain. Nevertheless, given our conservative estimate of the blazar luminosity density in Paper I, fixing it to the quasar luminosity density, we find that blazar heating raises the entropy substantially, starting around the epoch of He II reionization ($z \sim 3.5$) and by $z = 0$, the inclusion of blazar heating raises the entropy of the $\delta = 0$ fluid element to $K_e \approx (50 - 100) \text{ keV cm}^2$, approximately an order of magnitude larger than the case of photoheating alone.

⁷ In general, the efficiency for heating depends on the saturated, nonlinear damping rate of the maximally growing mode of the dissipating plasma instability, Γ_M , and the inverse Compton cooling rate, Γ_{IC} , and scales as $\Gamma_M/(\Gamma_M + \Gamma_{\text{IC}})$. We assume that the nonlinear damping rate is equal (or of order) the linear growth rate (see Section 3.5 of Paper I). For the linear growth rate of the oblique instability, this efficiency is close to unity for luminous blazars (Paper I). Numerical simulations for a mildly relativistic pair beam penetrating into a hot, dense background plasma suggest that a significant fraction ($\sim 20\%$) of the beam energy is heating the background plasma through the oblique instability before the two-stream plasma instability takes over, potentially dissipating another fraction of the beam kinetic energy (Bret et al. 2010).

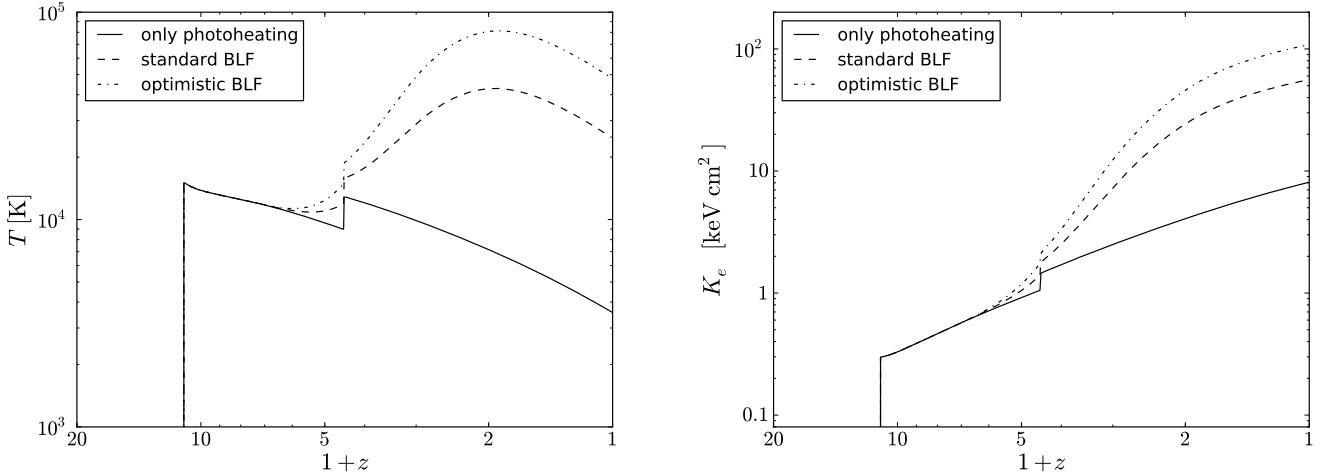


Figure 1. Thermal history (left) and entropy history (right) of a patch of mean density ($\delta = 0$) of the IGM. The solid curves are for pure photoheating with sudden reionization histories for H and He II at $z_{\text{reion}} = 10$ and $z_{\text{He II}} = 3.5$. The effect of the “loss of memory” in the thermal history of the IGM is apparent in the entropy as the pure photoheating curve asymptotes to $K_e \approx 8 \text{ keV cm}^2$ at $z = 0$. The dashed (dash-dotted) lines show the evolution for the standard (optimistic) blazar heating model that employs the blazar luminosity density, i.e., using the redshift evolution of the quasar luminosity density from Hopkins et al. (2007) and are normalized to the local heating rate. At $z = 0$, those differ by a factor of two and the standard one has been estimated from the observed number of blazars, employing a conservative estimate for the incompleteness correction, while the optimistic one naturally matches the inverted temperature–density relation in the Ly α data (Bolton et al. 2008; Viel et al. 2009; Puchwein et al. 2011).

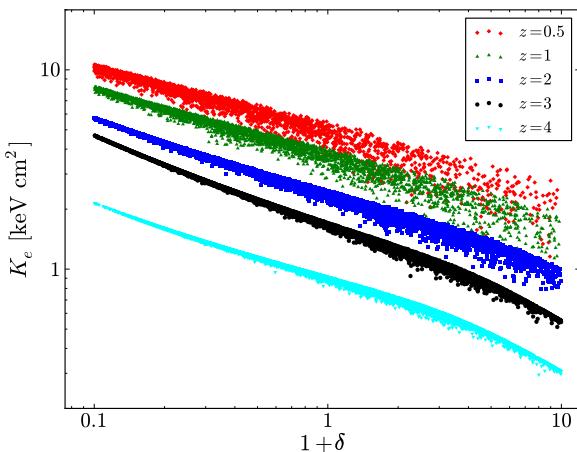


Figure 2. Entropy-density relation (K_e versus $1+\delta$) for $z = 0.5$ (red diamonds), 1 (green triangles), 2 (blue squares), 3 (black circles), and 4 (cyan inverted triangles) for a pure photoheating model.

As discussed in Paper II, the magnitude of the temperature enhancement due to blazar heating is density dependent: while the photoheating rate depends linearly on density, blazar heating is independent of density and dominates photoheating for regions with $\delta \lesssim 10$. Thus, blazar heating is expected to have the largest effects in voids. In Figures 2 and 3 we show the entropy-density relation (K_e vs. $1+\delta$) for a variety of redshifts, ranging from $z = 0.5$ to $z = 4$. For comparison, Figure 2 shows the case of photoheating only: the entropy-density relation reflects the effect of photoionization which deposits a uniform energy per baryon during H and He II reionization implying $K \propto (1+\delta)^{-2/3}$. This effect is quickly erased due to the “loss of memory” effect as well as adiabatic cooling due to the Hubble expansion approaching an asymptotically constant entropy–density relation at late times.

Figure 3 shows the effect of TeV blazar heating on the entropy-density relation for our standard (top) and optimistic (bottom) model. As expected, blazars have the most dramatic effects upon K_e in low-density regions ($1+\delta \simeq 0.1$) due to the spatially uniform and IGM density independence of their volumetric heating rate. This *steepens* the entropy-density relation, causing it to approach an asymptotics of $K \propto (1+\delta)^{-5/3}$ in voids. In fact, blazar-induced entropies can exceed those due to photoionization by a factor of 125–250 in these low-density regions ($1+\delta \simeq 0.1$, see right-hand panels of Figure 3), reaching $K_e \simeq (1250\text{--}2500)\text{keVcm}^2$ by $z = 0.5$. For high density regions ($1+\delta \simeq 10$) the effect is still pronounced, increasing the IGM entropy by up to an order of magnitude, i.e., $K_e \approx (10\text{--}20)\text{keVcm}^2$ at $z = 0.5$.

Apart from increasing the mean entropy, blazar heating also increases the scatter in entropy at fixed density which is especially noticeable for $z = 0.5$. Fractionally, this scatter appears to be roughly twice as large when blazar heating is included, though in absolute terms the scatter is much larger for the blazar cases. This is because even in linear theory, the entropy (and temperature) of any patch of overdensity δ depends on its “collapse” history. Namely, whether it collapses like a Zel’dovich pancake, i.e., in one dimension first which implies a smaller entropy and temperature, or more spherically which produces a greater entropy and temperature, i.e., more akin to accretion onto a local overdensity.

The fractional scatter in $K_{\text{blazar}}/K_{\text{photo}}$ is much larger than that found in either entropy individually. The reason is that the scatter induced by blazar heating is anti-correlated with that induced by photoheating due to their very different dependencies upon IGM density. Patches that begin at low densities when blazar heating is ignored are biased toward lower entropies as a consequence of an extended period of slow recombination. When blazar heating is included, these same patches are more efficiently heated, and therefore biased toward higher entropies.

The redshift evolution of blazar heating introduces additional

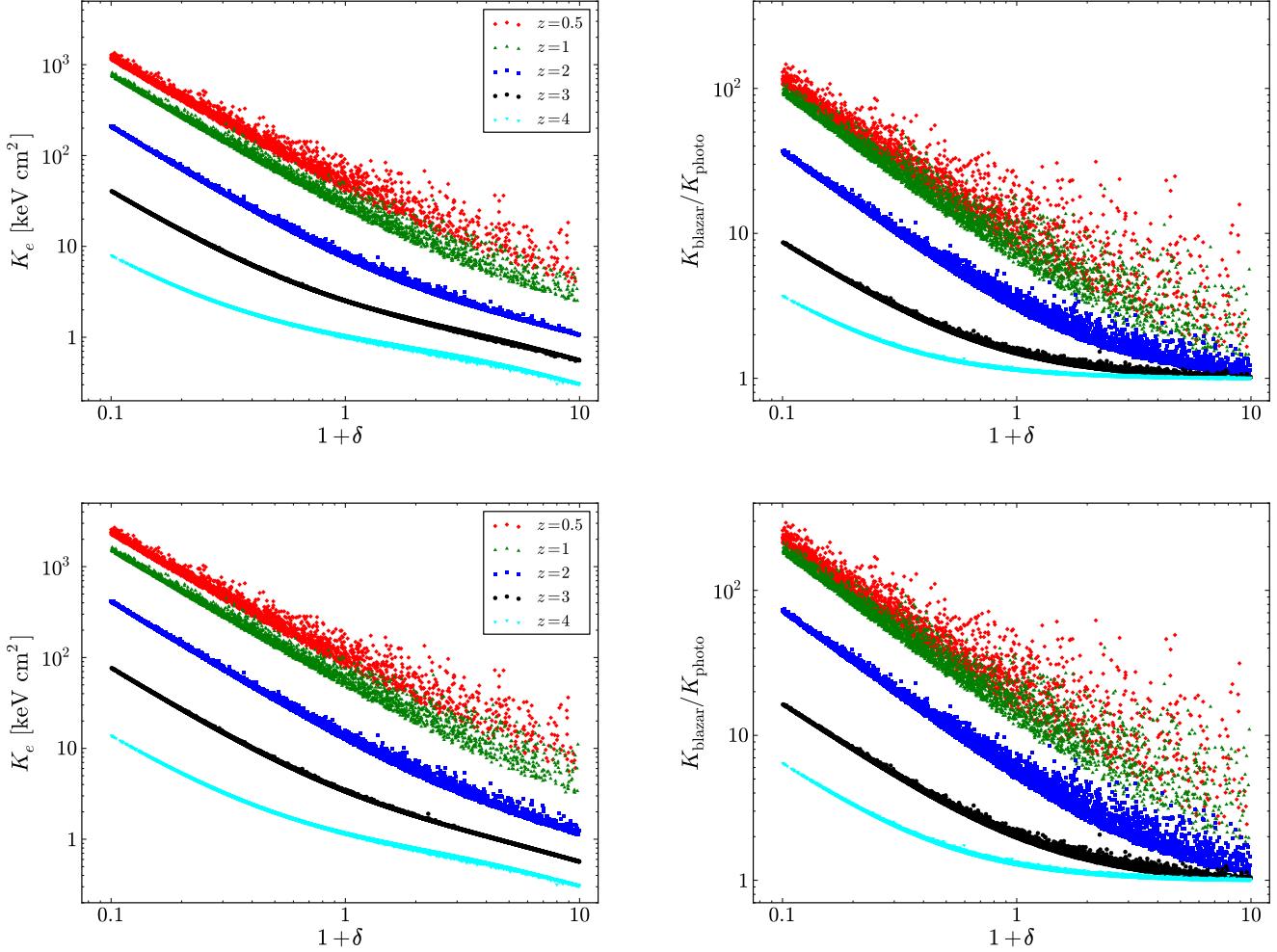


Figure 3. Entropy-density relation (K_e versus $1 + \delta$, left) for the standard (top) and optimistic (bottom) blazar heating model. Blazar heating increases the entropy over the pure photoheating case by a factor $K_{\text{blazar}}/K_{\text{photo}}$ which is shown versus $1 + \delta$ (right) again for the standard (top) and optimistic (bottom) blazar model. Shown are different redshifts $z = 0.5$ (red diamonds), 1 (green triangles), 2 (blue squares), 3 (black circles), and 4 (cyan inverted triangles).

stochasticity: patches at a given density sample a distribution of turnaround times and hence preheated entropy values. After turnaround, the gas is adiabatically compressed and moves on an adiabat (a line of constant K_e) to higher density with an entropy value that depends on the “collapse” time, and the steeper entropy-density relation generates the larger scatter at any $\delta \gtrsim 1$. Hence by preheating the universe, the first shells of gas that are able to collapse onto an object do not experience shock heating as they are only adiabatically compressed which results in an entropy floor of the object after formation. The later collapsing shells experience weaker shocks with a smaller Mach number due to their already elevated entropy level.

Associated with the large increase in the entropy of the IGM are a number of observational effects. In the following we discuss the impact blazar heating has upon the correlation between cluster X-ray luminosity and temperature, the entropy profiles and the CC/NCC bimodality of clusters and groups, and the SZ power spectrum. In each we note the unique effects imposed by the relatively recent nature of the entropy injection.

2.1. Implications for the X-ray emission of groups and clusters

Blazar heating raises the entropy floor for $z \lesssim 2$ with its peak contribution around $z \sim 1$ and acts on scales much larger than the turnaround region of clusters. The formation epoch of groups and clusters is roughly matched to the epoch of heating. As a consequence, early-forming, i.e., old groups may be little affected by blazar heating, had time to cool and develop an X-ray luminous core, potentially representing the class of CC groups/clusters. On the other hand, the blazar-heated IGM collapsing into late-forming, young groups will not have had sufficient time to cool. These groups will “remember” the elevated entropy floor as an extended core that substantially changes the initial conditions for their subsequent hierarchical evolution. However, because the heating occurs at late times, and thus after the first groups and clusters have begun to form, a larger energy input is required in comparison to the traditionally employed early preheating models.

For the observability of such a non-gravitational entropy floor and its implication on the thermal history of the ICM, it is important to consider its cooling timescale for typical entropy values of the IGM around $z \simeq 0.5$ and parameters typical of the

intra-group medium:

$$t_{\text{cool}} = \frac{3nkT}{2n_e n_H \Lambda(T, Z)} = 4.5 \text{ Gyr} \times \left(\frac{K_e}{75 \text{ keV cm}^2} \right)^{3/2} \times \left(\frac{kT}{1 \text{ keV}} \right)^{-1/2} \left(\frac{\Lambda(kT, Z)}{\Lambda(1 \text{ keV}, 0.3Z_\odot)} \right)^{-1}. \quad (2)$$

Here $n = 2.2n_H$, $\Lambda(T, Z)$ is the cooling function at a given T and metallicity, Z , (which is a relatively flat function of T near $kT = 1 \text{ keV}$ and $0.3Z_\odot$, Sutherland & Dopita 1993). A look-back time of 4.5 Gyr corresponds to a redshift of $z = 0.425$.

Instantaneous preheating by blazars (see Figure 1) is incapable of *directly* producing the highest entropies (up to 600 keV cm^2) that are presently observed in a few clusters (Cavagnolo et al. 2009). Comparing the cooling time to the dynamical time of a cluster/group of approximately ~ 1 Gyr or the typical timescales of significant turbulent pressure support of ~ 2 Gyr after a cluster merger which provides continuous heating throughout (Paul et al. 2011), we expect our heating to *indirectly* impact the entropy distribution through gravitational reprocessing of the blazar preheated core entropy, leading to a gravitational amplification of entropy that we now explain.

A larger central entropy, or equivalently, a lower central density⁸, of a merging cluster facilitates shock heating which implies an increase of the core entropy of the final object compared to that of a dense cooled core. CCs have a large inertia causing them to be rather resilient against shock heating, typically surviving the merger with only a marginally elevated entropy level which is then subject to fast radiative cooling. This effect gives rise to the well-known overcooling problem in cosmological simulations of galaxy clusters if cooling is not counteracted by any feedback process (e.g., Borgani & Kravtsov 2009).

In order to quantify the effect of gravitational reprocessing of a preheated entropy core, let us define a net entropy amplification factor through shock heating. This is easiest done by comparing two types of simulations of galaxy cluster formation: one with gravitational physics only and one where gravitational physics is supplemented by a preheating epoch of the IGM at mean density with a uniform entropy floor of K_{floor} which precedes the turnaround and formation of a galaxy cluster. In the first case, the core entropy structure is set by gravitational formation shocks leading to a central entropy of $K_{\text{grav},0}$. In the preheating case, the core entropy is raised to an elevated level after nonlinear evolution during virialization, $K_{\text{pre},0}$. Hence, the net entropy amplification factor is then defined as the ratio of the core entropy in the preheating case, $K_{\text{pre},0}$, to that obtained by adding the IGM entropy floor value to the entropy in gravitational heating case only, $K_{\text{floor}} + K_{\text{grav},0}$.

Non-radiative cosmological simulations of galaxy clusters and groups demonstrate that the net entropy amplification factor can reach values ranging from 3 to 5 for clusters and groups, respectively, in the high-entropy case with a preheating value of $K_{\text{floor}} = 100 \text{ keV cm}^2$ imposed at $z = 3$ (see Figures 1 and 2 in Borgani et al. 2005). For the low-entropy case with $K_{\text{floor}} = 25 \text{ keV cm}^2$, this amplification factor is only mildly reduced to 2–5 for clusters and groups, respectively. This entropy amplification factor seems to be reduced for radiative cluster simulations that include cooling and star formation, possibly due to

⁸ For a cluster with a given mass, a larger core entropy implies a lower gas density as the (constant) core temperature $kT \propto Kn^{2/3}$ reflects the virial value imparted by the cluster's potential depth.

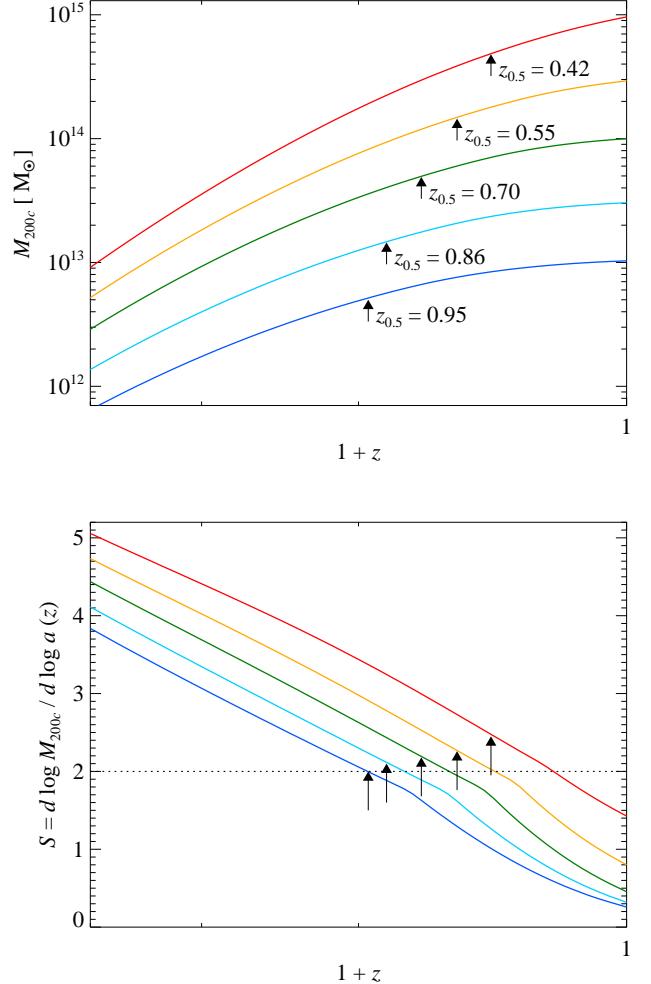


Figure 4. Median mass accretion histories (top) and mass accretion rates defined by the logarithmic slope $S = d \log M_{200c} / d \log a(z)$ (bottom) for five characteristic halo masses M_{200c} according to the model by Zhao et al. (2009). Arrows label the half-mass redshifts of the systems. At late times, the mass accretion rates on group scales start to flatten out which is a result of the change from the fast to the slow accretion mode. In contrast, clusters are still forming today as the half-mass redshift $z_{0.5}$ are still in the fast accretion regime characterized by $S > 2$.

the assumed early epoch of instantaneous entropy injection at $z = 3$ that facilitates cooling of the entropy floor thereafter (see Figures 3 and 4 in Borgani et al. 2005). Groups, however, still remain severely affected by an early epoch of preheating. Thus in combination with gravitational amplification, blazar heating could have an important effect on the subsequent thermodynamic evolution of late-forming groups and possibly the clusters they evolve into. This, however, is subject to nonlinear structure formation and needs to be carefully studied with cosmological hydrodynamic simulations containing a large sample of well-resolved galaxy clusters which sample the full distribution of formation redshifts.

The impact of the entropy injection due to blazar heating upon clusters is somewhat less clear. Clusters form in highly biased regions through the mergers of smaller, virialized systems such as groups. In most cases, these need to have already formed by $z = 1$ which can be inferred from the mass accretion history in Figure 4. This shows that on average, the most

massive group progenitor of a cluster with virial mass⁹ $M_{200c} = 10^{14} M_{\odot}$ crossed a mass threshold of $3 \times 10^{13} M_{\odot}$ before $z = 1$. At this time the entropy floor due to blazar heating was still smaller with typical values $K(\delta = 0, z = 1) \approx (25\text{--}50) \text{ keV cm}^2$ implying a smaller entropy core in groups at higher redshifts $z \gtrsim 1$. Hence, we naively may not expect blazar heating to have a large impact upon the X-ray luminosity and entropy profiles of clusters. However, this is not the case for two reasons.

First, the mass accretion rate, defined by the logarithmic slope $S = d \log M_{200c} / d \log a(z)$, is larger for more massive systems and at higher redshifts (where we have introduced the cosmic scale factor, $a \equiv 1/(1+z)$; see bottom panel of Figure 4). The mass accretion history of a halo observed at $z = 0$ grows on average as $M(z) = M_0 \exp(-\xi z)$ (Wechsler et al. 2002). The single free parameter in the model, ξ , can be related to a characteristic formation redshift $1+z_c = \xi/S$. Hence we find ourselves in the fast accretion regime if the mass accretion rate is larger than a characteristic value usually taken to be $S = 2$. In particular, the relative mass accretion rates increase from $z = 0$ to 2 by a factor 3 for clusters ($M_{200c} = 10^{15} M_{\odot}$) and 10 for groups ($M_{200c} = 10^{13} M_{\odot}$) (see also Gottlöber et al. 2001). A larger relative mass accretion rate implies that the gas in the core is heated by accretion processes at a faster rate. This means that a lower preheated core entropy value early-on can in principle be gravitationally processed at a substantially higher rate and, as a result of this, move onto a higher adiabat with a longer cooling time.

Second, blazars like AGNs follow the clustering bias of the matter density field and hence turn on first in highly biased regions, i.e., regions that evolve into clusters and super-clusters. At late times ($z \lesssim 3.5$) blazar heating is expected to be nearly spatially uniform, since more than a single blazar contribute significantly to the local heating rate of any given patch of the universe (Paper II). At much lower redshifts, the number of contributing sources grows dramatically, approaching 10^3 at $z = 0$, and thus today blazar heating is nearly homogeneous. Prior to $z \simeq 3.5$, however, blazar heating may exhibit $\sim 50\%$ fluctuations locally, due to the paucity of blazars in the early universe. By $z \simeq 6$ as much as 75% of the local heating can be due to a single object, implying large Poisson fluctuations in the heating rate. This results in a clustering bias at early times.

The rare first blazars are expected to appear in highly biased regions, corresponding to those that later evolve into groups and clusters, and therefore themselves be clustered. The likelihood of a patch of the universe being heated by such a blazar is then enhanced in the biased regions, where both the heating rate is larger (since the VHEGR density is larger) and the probability of being covered by a blazar in the cluster is larger.¹⁰ It is this early preheating in groups that may facilitate their evolution into NCC clusters (via the gravitational reprocessing of the high-entropy cores). Therefore, while we will focus

⁹ In this Section, we define the virial mass M_{200c} as the mass of a sphere enclosing a mean density that is 200 times the *critical* density of the universe.

¹⁰ In the limit of a few, highly clustered blazars, the VHEGR intensity declines exponentially from the clustering site for a spatially constant distribution of EBL photons, giving rise to an exponentially decreasing heating rate (on the scale of the VHEGR mean free path which is larger than the clustering length scale). However, if the distribution of EBL photons is also clustered and correlates spatially with the highly biased regions (which is expected in a hierarchically growing universe), the heating rate of those biased regions will be additionally enhanced in comparison to the void regions that have a low probability of being covered by a blazar. We stress that the effect of spatially and temporarily inhomogeneous blazar heating due to the clustering bias is expected to be absent at late times ($z \lesssim 3.5$) due to the frequent occurrence of blazars.

upon group/small-cluster mass scale in our study of the entropy structure immediately after group/cluster formation, gravitational reprocessing combined with the bias in the early blazars suggest that these effects may be important for more massive clusters as well, propagating the effect of blazar heating up the mass hierarchy.

The formation time of groups and clusters determines the blazar contribution to their entropy, inducing an intrinsic scatter in the $L_x - T$ relation associated with the distribution of collapse histories. As a result, we would expect to find systematically higher entropies in younger groups and possibly clusters provided gravitational reprocessing was effective. There is some evidence that this is the case. Optically selected, and therefore young, group and cluster samples have on average lower X-ray surface brightness and smaller gas mass fractions compared to X-ray-selected samples (Mahdavi et al. 2000; Hicks et al. 2008; Dai et al. 2010). Optically bright, small to moderately massive clusters ($kT > 4 \text{ keV}$) at redshifts $z = 0.6\text{--}1.1$ are under-luminous in X-rays for a given T , which implies a reduced gas density and by extension an enhanced core entropy by roughly a factor of two (Hicks et al. 2008). Similarly, an X-ray stacking analysis of *ROSAT* data based on optically selected groups at low redshift ($0.5 \text{ keV} < kT < 2 \text{ keV}$) from the Two Micron All Sky Survey catalog finds systematically lower gas mass fractions than expected, f_{gas} , (within an over-density of 500 times the critical density of the universe) and flatter temperature profiles (Dai et al. 2010)¹¹, again implying larger entropies. Taken at face value, this is in conflict with careful X-ray studies using *Chandra* data (Sun et al. 2009) if both samples are believed to represent the same underlying distribution. Alternatively, this is perfectly consistent if the entropy of clusters and groups varies with formation time, with the peak entropy injection occurring near $z \sim 1$. Both are well matched to the properties of AGN feedback generally, and blazar heating especially.

2.2. Thermodynamic structure of galaxy groups and clusters

The large injections of entropy necessarily influence the formation of cosmic structure. While the incorporation of an instantaneous entropy floor in numerical simulations does reproduce the $L_x - T$ scaling relations well, it is ad hoc and fails to produce the observed entropy profiles in a subset of X-ray luminous, cool core groups with steep entropy profiles and low values of the central entropy (see Section 1.1). Thus, it is clear that the effects of heating the ICM are more complicated than the introduction of a constant, global entropy floor (McCarthy et al. 2008; Fang & Haiman 2008). Here we explore whether blazar heating may be capable of producing the observed entropy profiles as well as some of the features of the distribution of large-scale structures observed.

While large-scale numerical simulations are required to study the entropy evolution of groups and clusters in detail, we can use the conservative property of entropy to estimate the effect of blazar heating on the thermodynamics of galaxy groups and (to a lesser extent) clusters. Smooth accretion, in which cold gas enters the cluster through a spherically symmetric accretion shock, results in self-similar entropy profiles

¹¹ We caution the reader that such a stacking analysis could have potential biases: the stacked X-ray spectrum in a given richness bin might be dominated by a few hot systems and fitting an average spectrum could then bias the temperatures and hydrostatic masses high and hence the gas fractions low. Edington bias from more numerous smaller systems can additionally lower the average gas fraction at a given optical richness. Before drawing far reaching conclusions, careful mock analyses are needed to confirm these results.

(Voit 2005). Characteristic values for the physical parameters are given by their average values within the virial radius, R_{200} (that we define as the radius of a sphere enclosing a mean density that is 200 times the *critical* density of the universe):

$$\begin{aligned} n_{e,200} &= 200 x_e X_H f_b \frac{\rho_{\text{cr}}}{m_p} = 1.7 \times 10^{-4} E^2(z) \text{ cm}^{-3} \\ kT_{200} &= \frac{GM_{200c} \mu m_p}{2R_{200}} = \frac{\mu m_p}{2} [10 GH(z) M_{200c}]^{2/3} \\ &= 1 \left(\frac{M_{200c} E(z)}{6 \times 10^{13} M_\odot} \right)^{2/3} \text{ keV} \quad (3) \\ K_{e,200} &= \frac{kT_{200}}{n_{e,200}^{2/3}} = \frac{\mu m_p^{5/3}}{2} \left(\frac{4\pi}{15} \frac{G^2 M_{200c}}{(1+X_H) f_b H(z)} \right)^{2/3} \\ &= 326 \left(\frac{M_{200c} E^{-1}(z)}{6 \times 10^{13} M_\odot} \right)^{2/3} \text{ keV cm}^2, \end{aligned}$$

where $f_b = \Omega_b/\Omega_m$ is the universal baryon fraction, the electron fraction is defined as the ratio of electron and hydrogen number densities, $x_e = n_e/n_H = (X_H+1)/(2X_H) = 1.158$, in which we assumed a fully ionized fluid with a primordial hydrogen mass fraction $X_H = 0.76$, and $\rho_{\text{cr}} = \rho_{\text{cr}}(z) = 3H^2(z)/(8\pi G)$ is the critical mass density, in which the Hubble function $H(z)$ is given by

$$\frac{H^2(z)}{H_0^2} = E^2(z) = (1+z)^3 \Omega_m + (1+z)^2 (1 - \Omega_m - \Omega_\Lambda) + \Omega_\Lambda. \quad (4)$$

Of particular importance here is that the entropy scales as $K_e \propto r^{1.1}$ (Voit 2005). The predictions of this simple model agree well with numerical simulations (Tozzi & Norman 2001) and the entropy profiles in the outer regions of clusters inferred from recent X-ray observations by *Chandra* and *XMM-Newton* (Cavagnolo et al. 2009; Pratt et al. 2010). Within the smooth accretion model the stratified profile is built from the inside out, with later accreted shells containing larger entropy due to gravitational heating at accretion shocks. Thus, if radiative cooling can be neglected, the entropy distribution of the gas in the cluster core reflects the entropy of the IGM immediately prior to the initial collapse of the group/cluster. This assumption is reasonable as long as the radiative cooling time (Equation (2)) is longer than the time interval between successive mergers, or equivalently the mass accretion timescale (see discussion in Sect. 2.1).

In practice, the smooth-accretion picture is over-simplified. Most of the ultimately accreted gas is not smoothly distributed, but rather contained within virialized substructures, and therefore has already been shock heated, fundamentally altering the way in which the entropy profile is generated. This results in a more complex morphology of the dissipating structure-formation shocks, which exhibit a rich network of shock fronts at which the gravitational energy of the gas is dissipated (Miniati et al. 2000; Ryu et al. 2003; Pfrommer et al. 2006). Nevertheless, despite this apparent chaos, non-radiative galaxy cluster simulations that have sufficiently high resolution find approximately self-similar entropy, density, and temperature structures outside of the core region, independent of the numerical method used (Frenk et al. 1999; Voit et al. 2005), yielding a universal entropy profile of

$$K_e(r,z) = 1.45 K_{e,200}(z) (r/R_{200})^{1.2}. \quad (5)$$

The reason for this is that the low entropy/high density gas ar-

ranges itself such that it finds itself at the bottom of the gravitational potential and eventually mixes with the surrounding halo gas as can be seen by rewriting the hydrostatic equation,

$$\frac{dP}{dr} = -\frac{GM(< r)}{r^2} \left(\frac{P}{K} \right)^{3/5}. \quad (6)$$

Magnetic draping of cluster fields during gravitational settling provides a thermal insulation of these low-entropy parcels (Lyutikov 2006; Dursi & Pfrommer 2008; Pfrommer & Dursi 2010), and hence this settling occurs adiabatically, producing a stratified core in which the hot, high-entropy halo gas remains at large radii.

If the universe is preheated, the central entropy of a newly formed group or cluster is replaced with a flattened core. Initially, the entropy of the core is set by that of the IGM at the time the second gas shell is accreted, forming an accretion shock and adiabatically compressing the gas in the first. Employing only heating by blazars and structure formation shocks, we show the resulting entropy profiles for groups of a variety of virial masses and formation redshifts in Figure 5, exploring the distribution around the median object mass. These are based upon Equations (3) and (5) as well as the floor values implied by Figure 1. Interestingly, the blazar-induced entropy floor at $z = 0.5$ is comparable to the gravitationally established entropies in groups of $3 \times 10^{13} M_\odot$ at radii $(0.2\text{--}0.4) R_{200}$ for our standard and optimistic blazar models, respectively. The implied core sizes are similar to observationally accessible radii of $R_{2500} \sim 0.3 R_{200}$ of optically selected clusters (see Figure 1 of Pratt et al. 2010). In our picture, these correspond to young clusters with recent formation times.

Thus far we have implicitly assumed the instantaneous formation approximation, i.e., that we may identify a particular redshift at which to calculate $K_{e,200}$ and the core entropy level due to blazar heating. This is naturally identified with the redshift at which the group/cluster has assembled half of its mass, after which strong structure-formation shocks develop. Figure 4 shows the accretion histories for groups/clusters in the mass range of relevance here, noting explicitly the half-mass redshifts for a variety of accretion histories. Typically, these occur after $z \simeq 1$, by which time blazar heating has had an opportunity to inject significant amounts of entropy into the IGM, implying that the TeV blazars can have a substantial impact upon the structure of groups/clusters in practice.¹² However, we note that median mass accretion histories are too simplified to assess the impact of blazar heating upon group/cluster entropy profiles in detail since they do not address the radial redistribution of accreted gas. Furthermore, we do not attempt to address the generation of entropy by gravitational reprocessing during mergers of later accreted material in Figure 5. After this explicit demonstration of the impact of a redshift-dependent entropy floor on cluster entropy profiles, we turn to the implications of these for the cluster population as a whole.

2.3. Implications for the bimodality of core entropy values

The thermodynamic properties of clusters in their centers show a clear bimodality, which is traditionally separated into

¹² The half-mass redshifts depend on the adopted definition for the halo mass (and of course cosmology which we fix here for simplicity). While we choose to use M_{200c} in this section, we note that the half-mass redshift are somewhat smaller when adopting the definition for the virial mass with an overdensity that varies with redshift (Bryan & Norman 1998). In this case, we obtain $z_{0.5} = \{0.85, 0.74, 0.62, 0.48, 0.36\}$ for our halos, $M_{\text{vir}} = \{1.3, 3.8, 13, 38, 130\} \times 10^{13} M_\odot$.

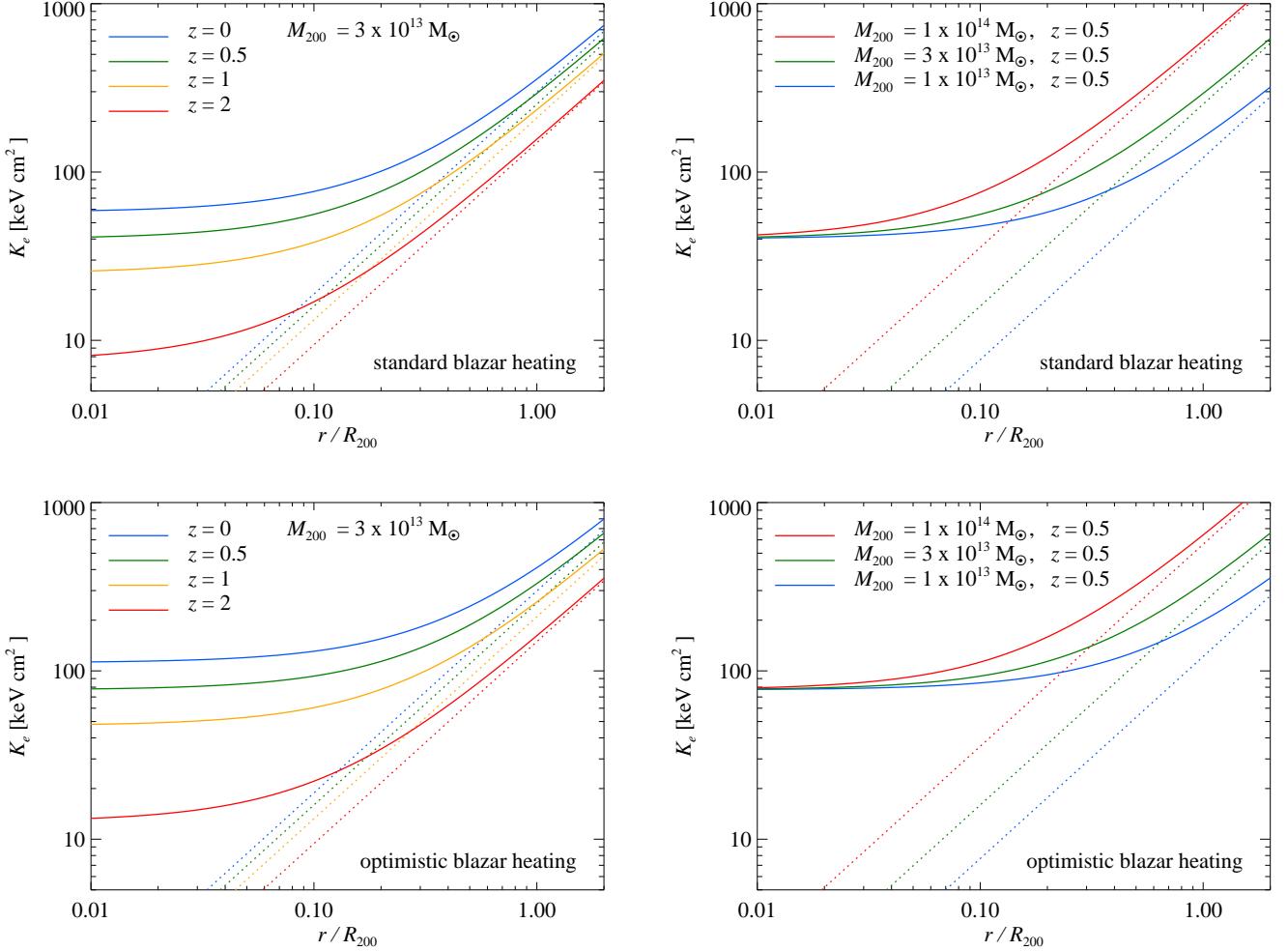


Figure 5. Radial profile of the entropy, $K_e = kT/n_e^{2/3}$, in galaxy groups/clusters for varying formation redshifts (left) and cluster masses (right) when our standard (top) and optimistic (bottom) blazar heating models are employed. While the outer profile results from dissipating gravitational energy in accretion shocks (dotted), the entropy core immediately after formation is set by the redshift-dependent blazar heating. Note that only late-forming groups ($z \lesssim 1$) are directly affected by this homogeneous but redshift-dependent preheating mechanism. If the ratio of cooling time, t_{cool} to the time period to the successive merger t_{merger} , is smaller than unity, the group will radiate away the energy associated with the elevated entropy core and evolve into a CC. Alternatively, if $t_{\text{cool}} > t_{\text{merger}}$, merger shocks can gravitationally reprocess the entropy cores and amplify them (not shown here). Potentially those groups may evolve into NCC systems.

two classes: CC and NCC clusters. The former are defined to have temperature profiles which decline significantly toward the center whereas the central temperature distribution of the latter remains constant and often correlates with merger events. As a result, the distribution of core entropy also appears to be bimodal in clusters, with the CC population peaking at $K_{e,0} \sim 15 \text{ keV cm}^2$ and NCCs at $K_{e,0} \sim 150 \text{ keV cm}^2$ separated by a gap between $K_{e,0} \sim (30\text{--}50) \text{ keV cm}^2$. Roughly half of the entire population of galaxy clusters in the *Chandra* archival sample show cooling times that are longer than 2 Gyr and have high-entropy cores typical of NCCs (Cavagnolo et al. 2009). This CC/NCC bimodality appears to be real and not due to archival bias as a complementary approach has shown with a statistical sample (Sanderson et al. 2009).

Before we discuss how blazars may impact on the relative abundance of cluster populations, we will shortly review how the currently favored hypothesis of AGN feedback compares to data. The possibility that AGN feedback can raise the core entropy $K_{e,0}$ to values representative of NCCs on the buoyancy timescale (e.g., Guo & Oh 2009) is not supported by observa-

tions. This is explicitly shown in Figure 6 which correlates $K_{e,0}$ with the volume work done by the expanding bubbles, E_{cav} , and with the cavity power, P_{cav} , in systems that show X-ray cavities inflated by AGN bubbles. The total energy required to inflate a cavity is equal to its enthalpy, given by

$$E_{\text{cav}} = \frac{\gamma}{\gamma - 1} PV_{\text{tot}} = 4PV_{\text{tot}} \quad (7)$$

assuming a relativistic equation of state $\gamma = 4/3$ within the bubbles. The cavity power is estimated using $P_{\text{cav}} = E_{\text{cav}}/t_{\text{buoyancy}}$ (Bîrzan et al. 2004; Rafferty et al. 2006).¹³ We compare the energy used to inflate the cavities to the gas binding energy of the

¹³ The buoyancy timescale could be either an overestimate of the true age since the cavity is expected to move outward supersonically during the early, momentum-dominated phase of the jet. It could also be an underestimate of the true age since magnetic draping provides an additional drag force slowing down the rise of the bubble (Dursi & Pfrommer 2008).

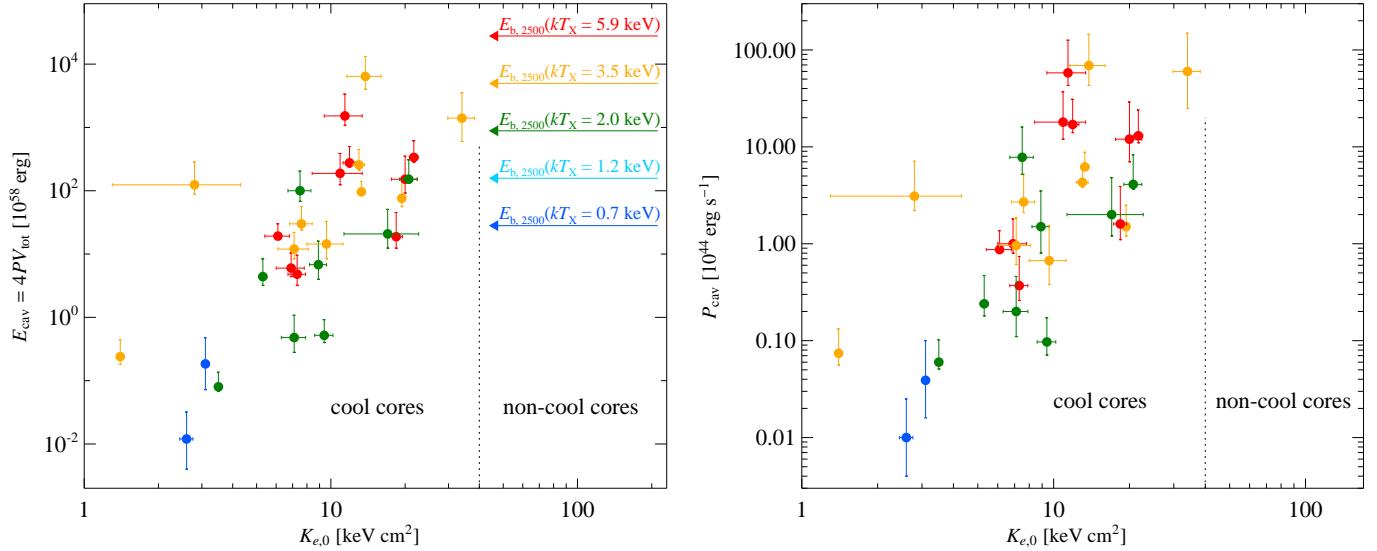


Figure 6. Correlating the observed core entropy values of clusters, $K_{e,0}$, (taken from the gradient extrapolation method of Cavagnolo et al. 2009) with the AGN cavity energy E_{cav} , as inferred from the volume work done by the expanding bubbles (left), and with the cavity power $P_{\text{cav}} = E_{\text{cav}}/t_{\text{buoyancy}}$ (right, from Rafferty et al. 2006). Color coding reflects average X-ray temperatures, the lower limit of each color bin is labeled in the upper right of the left panel. Arrows denote the gas binding energy contained within a spherical region of radius $R_{2500} \simeq R_{200}/3$. These very powerful AGN outbursts with energies up to 10^{62} erg and powers of 10^{46} erg s $^{-1}$ are in some cases energetically capable of unbinding the gas in the core regions. On the buoyancy timescale, no AGN outburst transforms a CC to an NCC cluster. This is apparent from the low core entropy values (with a median $K_{e,0} = 15$ keV cm 2) of typical CC clusters.

core region within a spherical region of radius $R_{2500} \simeq R_{200}/3$,

$$E_{b,2500} = f_{\text{gas},2500} \frac{GM_{2500}^2}{2R_{2500}} = \frac{f_{\text{gas},2500}}{2} M_{2500}^{5/3} [10 \text{ GHz}(z)]^{2/3} \quad (8)$$

$$\simeq 1 \times 10^{60} \text{ erg} \left(\frac{kT_X}{1 \text{ keV}} \right)^{3.23}$$

at $z = 0$. Here we use the phenomenological scalings obtained by X-ray observations of $h_{70} M_{2500} = M_5 (kT_X/5 \text{ keV})^{1.64 \pm 0.06}$, with $M_5 = (2.5 \pm 0.1) \times 10^{14} h_{70}^{-1} M_\odot$ (Vikhlinin et al. 2006) and $f_{\text{gas},2500} = (0.0347 \pm 0.0016) (kT_X/1 \text{ keV})^{0.509 \pm 0.034}$ (Sun et al. 2009).

P_{cav} or E_{cav} both measure the energy and power that is in principle available for heating the ICM, if it can be efficiently tapped by some process.¹⁴ However, as shown in Figure 6, even the most energetic and powerful AGN outbursts, with $E_{\text{cav}} \sim 10^{62}$ erg and $P_{\text{cav}} \sim 10^{46}$ erg s $^{-1}$, e.g., MS0735+7421 and Zwicky 2701, which are energetically capable of unbinding the gas in the core regions, are unable to disrupt the CC and transform the cluster into an NCC state on a buoyancy timescale. This is apparent from the low core entropy values (with a median $K_{e,0} = 15$ keV cm 2) of typical CC clusters. This is quantified with a linear Pearson correlation coefficient of 0.71 correlating $\log K_{e,0}$ with $\log P_{\text{cav}}$ or $\log E_{\text{cav}}$, respectively. This proves that the PdV work done by these expanding cavities is transferred inefficiently to the surrounding medium, i.e., the ICM

¹⁴ We note that of this $4PV_{\text{tot}}$, only PV_{tot} is available in form of mechanical energy while the internal energy $U = 3PV_{\text{tot}}$ is presumably still stored within the bubbles. What fraction of this internal energy is eventually thermalized, and thus can potentially contribute to unbinding the cluster gas and/or raising the core entropy, is not a priori clear. If the energy is stored in cosmic rays or magnetic fields, it may be transferred to the thermal pool via cosmic ray Alfvén-wave heating (Kulsrud & Pearce 1969) or magnetic reconnection, respectively. Nevertheless, the uncertainty induced by this is small in comparison to the many orders of magnitude increase in cavity energy required to explain the non-cool core clusters.

entropy produced by AGN inflated bubbles is far less than the virial value of $K_e \sim 540 \text{ keV cm}^2 \times (M_{200c}/M_{14})^{2/3}$, but enough to arrest overcooling.

On the other hand, there is a strong anticorrelation between the radio power of the brightest cluster galaxy and $K_{e,0}$ for nearby clusters ($z < 0.2$), implying that bright radio emission is preferentially “on” for $K_{e,0} \lesssim 40 \text{ keV cm}^2$ (see Figure 2 in Cavagnolo et al. 2008). While AGN feedback seems to be unable to transform a CC to an NCC cluster (on a buoyancy timescale), it appears to be critical in stabilizing the thermal atmospheres from entering a cooling catastrophe and collapsing. In principle the impact of AGN-induced turbulence on heat transport (conductively and advectively) could result in a CC to NCC metamorphosis on a much longer ($> \text{Gyr}$) timescale (Parrish et al. 2010; Ruszkowski & Oh 2010). This is because the temperature difference between the maximum of the temperature profiles to the cold center is at most a factor of three (Vikhlinin et al. 2006), implying that heat transport could initially increase the central entropy by a similar factor as $K_{e,0} \propto kT$. In the absence of radiative cooling, the associated pressure enhancement would adiabatically expand and thereby cool the gas, hence restoring the temperature gradient. Sustained conduction and advection could further increase the central entropy. However, the timescale for such a process is long in comparison to radiative cooling timescales ($\lesssim 1 \text{ Gyr}$) as well as cluster assembling and merging timescales which questions the possibility of this mechanism to boost the central entropies to values of $K_{e,0} \simeq 600 \text{ keV cm}^2$ which are at the tail end of the distribution (Cavagnolo et al. 2009).

Because blazar heating does not inject large amounts of entropy at $z \gtrsim 2$, and blazars do not efficiently heat high density regions ($1 + \delta > 10$), objects that have already collapsed by that time or shortly thereafter will not be significantly affected by blazars (barring the effect of clustering bias at early redshifts). In our scenario, these early-forming groups evolve into

CC systems at the present epoch that potentially need to be stabilized by a self-regulated feedback process, e.g., provided by the radio mode of AGNs. However, the subset of groups that forms after $z \simeq 1$ can be severely affected. If such a group is viewed shortly after its formation, it should still exhibit the elevated core entropies associated with the prior blazar heating. If such a late-forming group has a cooling timescale long in comparison to the interval between cluster/group mergers, merger shocks can gravitationally reprocess the entropy cores and amplify them by a factor of up to five (Borgani et al. 2005, see also Section 2.1). These late-forming groups would then evolve into NCC systems.¹⁵ For sufficiently late-forming clusters that experience a series of fast successive merger events and, hence, avoid substantial cooling phases, we expect that gravitational reprocessing should boost the central core entropy in the most extreme case from $\sim 100 \text{ keV cm}^2$ to $\sim 600 \text{ keV cm}^2$, allowing for reprocessing of the blazar-heated entropy floor due to gravitational heating. Typically, however, blazar heated entropies at the time of turnaround for late-forming groups/clusters are $\sim 50 \text{ keV cm}^2$. Allowing for modest cooling periods in between mergers should yield smaller median core entropy values of $\sim (100 - 150) \text{ keV cm}^2$ after accounting for gravitational reprocessing. These estimates compare favorably with observed values of NCC clusters (Cavagnolo et al. 2009).

We point out that it is very natural for systems with a unimodal distribution in core entropy values following group formation to evolve into a bimodal distribution today as consequence of both the cooling instability and the gravitational reprocessing of (temporally increasing) elevated entropy cores. The observed CC and NCC cluster populations are centered upon the two attractor solutions that a galaxy cluster can evolve into. In the case of CC clusters, the core evolution is driven by the well-known overcooling problem, encountered in cosmological simulations of galaxy clusters. Below a critical core entropy, purely hydrodynamical mergers are incapable of disrupting a compact CC systems and transforming it into an NCC object (Poole et al. 2008). For NCC clusters, the core evolution is driven by the rapid (in comparison to t_{cool} , which effectively sets the critical core entropy value within a given epoch) succession of mergers, which has been demonstrated to further elevate the core entropy values substantially (Borgani et al. 2005). Interestingly, this solution is not a runaway solution; instead, the gravitational bootstrapping of blazar preheated entropy adjusts to the system size. Hence, in this picture, the core entropy should never be able to exceed the entropy at the virial radius (according to virial arguments) and most likely reach only values that are a fraction of that (at least for cluster systems) due to radiative cooling and the modest blazar preheated entropy values in comparison to early preheating models.

To conclude, we demonstrate explicitly that the core entropy values in CC clusters have a very weak correlation with the mechanical energy and power of X-ray cavities inflated by AGNs. Hence, PdV work done by these expanding cavities is transferred inefficiently to the surrounding medium. This strongly

¹⁵ Using *Chandra* observations, there has been a claim that incidence rate of cool core clusters at redshifts $z > 0.5$ is much smaller than their fraction at low redshifts (Vikhlinin et al. 2007). We note that the development of the characteristic cuspy X-ray brightness profiles for cool cores at $z = 0$ requires the cooling time to be considerably shorter than the formation time—a criterion that is often not fulfilled at the (high) redshift in question. Second, in order to construct their sample, Vikhlinin et al. (2007) discarded regions with AGN emission which can be correlated with cool cores, potentially biasing their absolute cool core rates low. Hence, this observation is not in contradiction with our proposed scenario.

suggests that while AGN feedback seems to be critical in stabilizing CC systems, it cannot transform CC into NCC systems (at least on the buoyancy timescale). With this evidence it seems even more pressing to pursue alternative solutions such as the presented blazar heating scenario in combination with gravitational reprocessing that provides a plausible scenario for the observed CC/NCC bimodality. Future cosmological hydrodynamical simulations that include the effect of clustering bias of blazar heating are needed to study these considerations in greater detail.¹⁶

2.4. Implications for the Sunyaev-Zel'dovich power spectrum

The thermal SZ effect provides a direct probe of the gas properties of groups for $z \gtrsim 0.5$. The SZ effect arises from CMB photons that inverse Compton scatter off thermal electrons within the hot plasma in galaxy clusters and groups, producing a localized perturbation to the CMB spectrum (Sunyaev & Zel'dovich 1972; Sunyaev & Zeldovich 1980). The thermal SZ effect directly measures the thermal electron pressure in the gas and has the important property that its amplitude is independent of redshift. The pressure fluctuation spectrum of unresolved groups and clusters dominates the CMB power spectrum on angular scales smaller than $3'$ (corresponding to a multipole moment $\ell \simeq 3000$) and half of the SZ power spectrum signal at $\ell \simeq 3000$ comes from groups with $M_{500} < 2 \times 10^{14} M_{\odot}$ and $z > 0.5$ (Trac et al. 2011; Battaglia et al. 2011). At these scales, the SZ power spectrum depends on the square of the Fourier transform of the average pressure profile of clusters/groups; a more concentrated pressure profile implies more power, a smoother one less.

A population of groups/clusters with high core entropies, $K_0 = (50-100) \text{ keV cm}^2$, implies a smoother pressure core distribution than if the cores had cooled since formation and developed a more concentrated entropy profile. This is very similar to the effect of AGN feedback which injects entropy into the core of groups, smoothing out the resulting pressure profile (Battaglia et al. 2010). Adopting empirically motivated “universal” pressure profiles constrained by X-ray observations, the peak amplitude of the SZ power spectrum is reduced for NCC pressure profiles as compared to those of CC clusters (Efstathiou & Migliaccio 2011). Hence, we expect the effect of blazar heating to result in a suppression of the SZ power spectrum for scales $\ell \gtrsim 2000$, which probe the pressure profile mostly inside R_{500} . An abundant population of late-forming groups with high K_0 -values (that has not been taken into account in any numerical modeling of the SZ power spectrum so far, e.g., by Battaglia et al. 2010 or Trac et al. 2011) would furthermore reduce the thermal SZ power spectrum in comparison to these numerical approaches. This has potentially important observational consequences since angular scales around $3'$ are the sweet spot for current telescopes measuring the high- ℓ CMB angular power spectrum, e.g., the South Pole Telescope (SPT; Lueker et al. 2010; Shirokoff et al. 2011; Keisler et al. 2011; Vanderlinde et al. 2010) and the Atacama Cosmology Telescope (ACT; see, e.g., Fowler et al. 2010; Dunkley et al.

¹⁶ A detailed prediction of the core entropy distribution of clusters at a given mass (or temperature) depends on the merger history of all objects at that mass (to understand which number fraction of clusters was channeled into the cooling branch due to comparatively fast radiative cooling), the distribution of the departure times of the gas that ended up in the core from average densities (to address the exact magnitude of blazar heating for the population of clusters rather than individual systems), and quantifying the efficiency of gravitational re-processing for the entire population of groups and clusters (Borgani et al. 2005) only simulated a group and a cluster for different variants of physics).

2011; Marriage et al. 2011).

In addition to the astrophysical dependence, the amplitude of the SZ power spectrum also depends very sensitively on cosmological parameters that are responsible for the growth of structure, $C_\ell \propto \sigma_8^{7\text{--}9} (\Omega_b h)^2$, where the rms amplitude of the (linear) density power spectrum on cluster-mass scales is denoted by σ_8 . Hence, there is an interesting degeneracy in the amplitude (and shape) of the SZ power spectrum between the cosmological information (dominated by σ_8) and the astrophysical information contained in the average pressure profile. Currently, numerical models of the SZ power spectrum are consistent with the data at the 1σ level (Dunkley et al. 2011; Shirokoff et al. 2011). However, after allowing for a substantial signal from patchy reionization (Iliev et al. 2007, 2008), which boosts the kinetic SZ effect and hence the total SZ signal, the power predicted by these models becomes uncomfortably high. Suppression of power due to blazar heating represents a promising mechanism by which to reconcile the expected and observed SZ power spectrum.

Unfortunately, the impact of blazar heating is degenerate with energy injection from AGN feedback in the dense cores of groups at early times (where there are also very little constraints). Here, we sketch a promising idea on how to discriminate between the effects of AGN feedback and blazar heating on the SZ power spectrum. In particular, AGN feedback and blazar heating vary as a function of cluster mass and redshift. Their effects on the pressure profile of groups and clusters and, hence, their imprint on the SZ signal also vary as a function of cluster mass and redshift. By cross-correlating CMB maps (where foreground components and primary anisotropies have been subtracted) with deep optical redshift surveys that are separately binned in redshift and cluster mass, i.e., optical richness estimators, we can perform SZ tomography. Such tomographic SZ power spectra (similar to Figures 7 and 8 of Battaglia et al. 2011, for the case of AGN feedback) will enable us to derive a redshift and mass-dependent mean cluster pressure profile. Potentially, this could disentangle the different non-gravitational energy injections of AGN feedback and blazar heating in clusters. However, large cosmological hydrodynamical simulations are required to obtain detailed predictions of either process which shall be subject to future work.

3. STRUCTURE FORMATION AND DWARFS

The entropy injected into the IGM by blazars not only modifies the structures of groups and potentially clusters, but also has an observable effect upon cosmological structure formation. Heating the IGM produces higher IGM pressures, which in turn suppress the gravitational instability on sufficiently small scales. Thus, there is a characteristic length scale, and hence characteristic mass (M_C), below which objects will not form. The particular value of this critical mass depends upon how structures form in practice, and generally requires a fully nonlinear study of structure formation (e.g., that provided by large-scale numerical simulations). Nevertheless, we may estimate the relevant characteristic length and mass scales via linear perturbation theory.

A very rough idea of the impact of blazar heating may be derived from the Jeans wavenumber, obtained by balancing the sound crossing and free-fall timescales:

$$k_J(a) \equiv \frac{a}{c_s(a)} \sqrt{4\pi G \bar{\rho}(a)}. \quad (9)$$

where for convenience, we have introduced the cosmic scale

factor, $a \equiv 1/(1+z)$, $\bar{\rho}(a) \equiv \Omega_m(a) \rho_{\text{cr}}(a)$ is the mean total mass density of the universe, $c_s(a) \equiv \sqrt{5kT(a)/3\mu m_p}$ is the linear sound speed (in which T denotes the temperature at mean density and $\mu = 0.588$ denotes the mean molecular weight for a fully ionized medium of primordial element abundance). In a static background universe, perturbations on scales smaller than $2\pi a/k_J$ are stable, i.e., the ambient gas pressure is sufficient to counteract gravitational collapse. As a result, cooling, fragmentation, and star formation are suppressed within objects of mass less than the Jeans mass, M_J ,

$$M_J(a) \equiv \frac{4\pi}{3} \bar{\rho}(a) \left(\frac{2\pi a}{k_J(a)} \right)^3 = \frac{4\pi^{5/2}}{3} \frac{c_s^3(a)}{G^{3/2} \bar{\rho}^{1/2}(a)}. \quad (10)$$

The Jeans mass associated with a universe which has been heated by TeV blazars, $M_{J,\text{blazar}}$, is necessarily larger than one which has only undergone photoionization heating, $M_{J,\text{photo}}$ by

$$\frac{M_{J,\text{blazar}}}{M_{J,\text{photo}}} = \left(\frac{c_{s,\text{blazar}}}{c_{s,\text{photo}}} \right)^3 = \left(\frac{T_{\text{blazar}}}{T_{\text{photo}}} \right)^{3/2} \gtrsim \begin{cases} 18, & \text{stand. model,} \\ 50, & \text{opt. model,} \end{cases} \quad (11)$$

where we used a temperature ratio of $T_{\text{blazar}}/T_{\text{photo}} \gtrsim \{7, 14\}$ at the present epoch for the standard and optimistic model, respectively (see Figure 1 and the discussion surrounding the standard and optimistic blazar heating histories). That is, blazar heating increases the mass of the smallest collapsed objects by more than an order of magnitude with the exact value depending on the adopted numbers of blazars contributing to the heating.

While the Jeans mass provides a way to estimate the importance of blazar heating, M_J typically exceeds M_C by up to an order of magnitude because it neglects the Hubble expansion. That is, it fails to account for the time required for the pressure to influence the evolving gas distribution. A more rigorous treatment of linearized density perturbations in a baryon-dark matter fluid finds that the overdensity in dark matter, $\delta_d(t, k)$, and the overdensity in the baryon distribution, $\delta_b(t, k)$, both of which are functions of time and comoving wavenumber, are related by

$$\frac{\delta_b(t, k)}{\delta_d(t, k)} = 1 - \frac{k^2}{k_F^2} + \mathcal{O}(k^4), \quad (12)$$

for some k_F (Gnedin & Hui 1998). The “filtering scale”, associated with k_F , defines a size below which baryonic perturbations are smoothed despite the growth of background dark matter perturbations. The filtering wavenumber can be related to k_J by

$$\frac{1}{k_F^2(t)} = \frac{1}{D_+(t)} \int_0^t dt' a^2(t') \frac{\dot{D}_+(t') + 2H(t')\dot{D}_+(t')}{k_J^2(t')} \int_{t'}^t \frac{dt''}{a^2(t'')}, \quad (13)$$

where $D_+(t)$ is the linear growth function and is dependent upon the cosmology (Gnedin 2000). While k_F is related to k_J , at any time those can be very different since k_F is an integral over the past evolution of the Jeans scale, weighted by the appropriately scaled growth function.

The associated mass scale, defined in analogy with the Jeans mass, is

$$M_F(a) \equiv \frac{4\pi}{3} \bar{\rho}(a) \left(\frac{2\pi a}{k_F(a)} \right)^3, \quad (14)$$

(the details of how this is computed in practice are collected in Appendices A and B). To understand the physics underlying

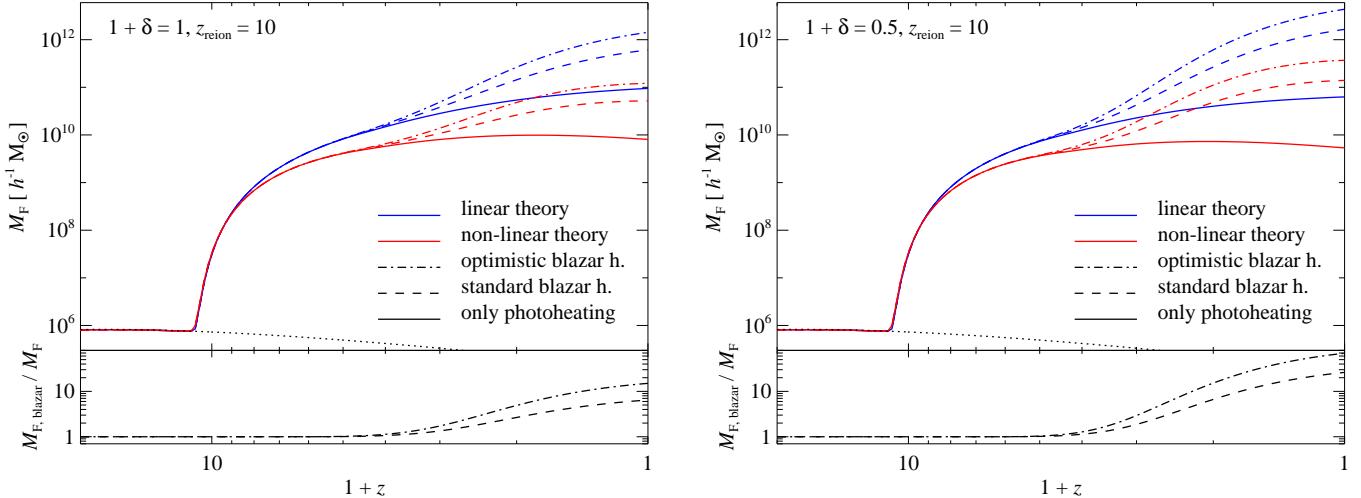


Figure 7. Redshift evolution of the filtering mass, M_F , for the cosmic mean density, $\delta = 0$ (left) and for a void with mean overdensity, $\delta = -0.5$, (right). We contrast M_F in the standard cosmology that employs only photoheating (solid) to the case of blazar heating in our standard model (dashed) and optimistic model (dash-dotted). In the bottom panels, we show the ratio of M_F in our respective blazar heating models to those without blazars. To estimate the effect of nonlinear structure formation on the filtering mass, we compare the linear theory M_F (blue) to the nonlinear theory M_F (red) where we used a correction function derived from hydrodynamic simulations by Okamoto et al. (2008). The dotted line represents the analytical solution for M_F which is only valid before reionization for a power-law temperature evolution.

the filtering mass, it is instructive to connect the filtering mass to the entropy of the IGM. To this end, we use the definition for the entropy K of Equation (1). As we show in Appendix A, the filtering scale $\lambda_F = 2\pi a/k_F$ is an integral over the entropy evolution,

$$\frac{1}{k_F^2(a)} = \frac{A_0}{D_+(a)} \int_0^a da' K(a') \frac{D_+(a')}{a'^3 E(a')} \int_{a'}^a \frac{da''}{a''^3 E(a'')}, \quad (15)$$

$$A_0 = \frac{5}{3} \left(\frac{3\Omega_m}{8\pi G H_0} \right)^{2/3}. \quad (16)$$

Here $E(a) = H(a)/H_0$ is the dimensionless Hubble function. Since $M_F \propto k_F^{-3}$, the linear filtering mass is also determined by the entropy evolution at mean density and appropriately scaled with the linear growth function, Hubble function, and cosmic scale factor. Since the integrand in Equation (15) is a positive definite function, the linear filtering mass increases monotonically with time.

We expect the strong increase of the entropy due to blazar heating to drive a comparably strong increase in M_F , though slightly delayed due to the effective weighting function in M_F . Physically, this implies that the entropy floor delivered by photo- or blazar heating at densities around the cosmic mean is then conserved by adiabatic compression during structure formation and directly translates to a linear filtering mass M_F . However, during this nonlinear process of structure formation, the entropy can be additionally augmented through dissipation of gravitational energy in cosmic structure-formation shocks or decreased through radiative cooling. The competition of these two processes determines which will dominate and in turn sets the critical mass M_C which is required for the condensation of gas into a dark matter halo and eventually the formation of a galaxy.

As with M_J , M_F typically exceeds the M_C observed in simulations by a substantial factor. Whether a halo can accrete gas is determined by the gas temperature (or equivalently entropy) at the virial radius which can only be obtained through hydrodynamic cosmological simulations. Estimated values for

this critical mass using cosmological simulations of nonlinear structure formation with only photoheating yield $M_C(z=0) = 6.5 \times 10^9 h^{-1} M_\odot$ (Hoeft et al. 2006; Okamoto et al. 2008)—one order of magnitude smaller than the linear analog M_F at $z=0$ (Okamoto et al. 2008, see also Figure 7). This discrepancy becomes less for increasing redshift leading to reasonably good agreement at high redshift ($z > 6$) between M_F and M_C . To model the nonlinear behavior of M_F , we can introduce a correction factor $C(z) = M_C(z)/M_F(z) = (1+z)^{1.1}/11.8$, where $M_F(z)$ has been modeled from the temperature evolution at mean density in the cosmological simulations and includes weakly nonlinear aspects of structure formation (Okamoto et al. 2008; Macciò et al. 2010). However, validating this approximation for M_C will ultimately require numerical simulations that include blazar heating.

Figure 7 shows M_F as a function of redshift for a void and the cosmic mean when various heating mechanisms are considered (which agrees for the standard cosmology with Figure 5 of Gnedin 2000). Note that for all of the thermal histories we considered, the resulting M_F nicely follows the expected analytical form over its full range of validity prior to recombination (which necessarily implies an Einstein-de Sitter universe, see the derivation in Appendix B). While blazar heating increases M_F in voids as early as $z \sim 3$, perturbations at the cosmic mean are affected slightly later ($z \sim 2.5$). The extra suppression of the formation of small galaxies due to blazar heating amounts to a mass suppression factor of $\sim 25 - 70$ for voids and $\sim 6.5 - 15$ for the cosmic mean today. The range indicates the uncertainty of the blazar heating models and the larger value belongs to the model that matches the inverted temperature-density relation in the Ly α forest found by Viel et al. (2009), i.e., the optimistic model. The resulting estimates for the nonlinear characteristic mass M_C are also shown in Figure 7 with red lines for our models with and without blazar heating. Apparently the increase of M_C due to blazar heating is able to counteract the suppression of M_F in nonlinear theory. We note that our estimate for M_C is slightly decreasing for $z < 1$ in the model that employs only photoheating. This decrease is an ar-

tifact of our linear theory M_F which by construction does not model nonlinear aspects of structure formation such as shell crossing or formation shocks that raise the temperature of some patches already at mean density.

We now turn to two outstanding problems in cosmological structure formation that the recent blazar heating may help to address: the missing satellite (Section 3.1) and void dwarf (Section 3.2) problems. Finally in Section 3.3 we discuss how blazar heating may naturally address some of the difficulties with WDM cosmologies of galaxy formation.

3.1. Dwarf satellites in the Milky Way

The heating due to blazars provides an additional mechanism to suppress the formation of dwarfs. Unlike photoionization models, which typically invoke the heating at reionization, blazar heating provides a well defined, time-dependent suppression mechanism, with the suppression rising dramatically after $z \sim 2$. This can be seen explicitly by the steep increase in filtering mass for these redshifts in Figure 7. In addition, due to the homogeneous nature of the heating, i.e., a constant volumetric rate that is independent of density, the heating from blazars suppresses structure formation most efficiently in the low-density regions responsible for late-forming dwarf halos due to their negative bias, where the energy deposited per baryon is larger.

In fact, the star formation histories of dwarf galaxies provide a strong constraint upon the magnitude of blazar heating: if dwarf galaxy formation is suppressed by high-energy gamma-ray emission from blazars, all dwarf star formation histories must begin prior to $z \simeq 2$, roughly the redshift at which blazar heating becomes significant. Over the past decade most of Local Group dwarfs have been observed with the *Hubble Space Telescope* shifting the focus from counting dwarfs to resolving the individual stellar population within these objects. Thus, it has become possible to construct detailed color-magnitude diagrams of every dwarf galaxy and therefore their detailed star formation histories (Dolphin et al. 2005; Holtzman et al. 2006; Orban et al. 2008). While the data shows a great variety of star formation histories—some continuous, some bursty, some truncated—they all have in common that they extend beyond a lookback time of 10 Gyr (corresponding to $z = 2$). There is no known case of a dwarf galaxy that formed its first set of stars after $z \simeq 2$. The fact that there exists an old stellar population in every dwarf in the Local Group provides a very important test that our model of blazar heating successfully passes.

The degree to which blazar heating can suppress the number of dwarf galaxies depends upon the redshifts at which they are typically formed. By combining high-resolution N -body simulations of the evolution of Galaxy-sized halo with semianalytic models of galaxy formation, Macciò & Fontanot (2010) and Macciò et al. (2010) have inferred the statistics, formation time,¹⁷ and accretion histories of the Milky Way satellites. Their findings are in good agreement with recent work by other groups (Koposov et al. 2009; Muñoz et al. 2009; Guo et al. 2010; Busha et al. 2010; Font et al. 2011) and thus can be regarded as representative for this model class. They find that the distribution of formation times is bimodal as a result of the suppression of hot gas accretion in low-mass halos

¹⁷ Macciò et al. (2010) defined the formation time of dwarfs as the redshift where the progenitors of dwarfs that end up within the Milky Way halo today exceeded a virial temperature of $T_{200} = 10^4$ K such that H I cooling of the gas becomes possible in these halos. That virial temperature corresponds to a mass threshold of $M_{200} \simeq 10^9 M_\odot$ at $z = 1$ and $10^8 M_\odot$ at $z = 10$.

due to the photoionization background; while 2/3 of today's satellites form at redshifts ranging from $3 < z < 12$, 1/3 of all satellites form late at $z < 3$ with most of them at $z < 1.5$. Because their formation is marked by the first time H I cooling can effectively form the first stars, their stellar populations are necessarily younger than their formation time of $z < 1.5$ in direct conflict with observed ages of the stellar populations in all of the Local Group dwarfs¹⁸ (Dolphin et al. 2005; Holtzman et al. 2006; Orban et al. 2008). The model of Macciò & Fontanot (2010) that successfully reproduced the satellite luminosity function of the Milky Way assumed the linear theory filtering mass (Gnedin 2000, see solid blue line in Figure 7) which was shown to significantly overproduce the characteristic halo mass scale below which baryons cannot condense and form stars (Okamoto et al. 2008, see solid red line in Figure 7). Adopting their less efficient nonlinear filtering mass formalism, the number of satellite galaxies with intermediate magnitudes $M_V \sim 10$ increases and creates a bump in the luminosity function which is clearly inconsistent with the data (Macciò et al. 2010).

Inspection of nonlinear estimates for the filtering mass M_F in Figure 7 shows that heating from blazars could prevent the condensation of baryons in halos of masses $10^{10} M_\odot$ at redshifts $z \sim 2$ and likely up to $10^{11} M_\odot$ at $z = 0$. At mean density, these nonlinear estimates for M_F correspond to the original linear values of Gnedin (2000) which have been shown to yield the observed dwarf satellite abundances and luminosity functions (Somerville 2002; Macciò et al. 2010). At the same time, the strongly rising M_F in the blazar heating models after $z \sim 2$ is able to suppress the population of late-forming dwarfs (Macciò & Fontanot 2010) that are in conflict with measured star formation histories of dwarf galaxies. Thus, not only can blazar heating potentially play a significant role in explaining the observed abundances of dwarf galaxies but also their old star formation histories.

Finally, we argue that the redshift evolution of the blazar heating rate should manifest itself as stochasticity in the satellite luminosity function at fixed host halo mass. Host halos show a distribution of formation times (e.g., Wechsler et al. 2002). This distribution is inherited by the host's satellite dwarfs which on average form earlier than the host. As demonstrated in Figure 1, blazar heating implies an entropy floor that dramatically increases after $z \sim 2$. Hence, the distribution of formation times of satellite dwarfs at fixed host halo mass results in a distribution of pre-collapsed entropy of these dwarf halos. As a result, this leaves early-forming dwarfs relatively unchanged but suppresses the baryon fraction or even the formation of late-forming dwarfs. This results in different cooling histories for different dwarfs, which, in turn, might modify the stellar content that condenses out in these systems. Hence, blazar heating provides an apparently substantial, physically motivated scatter in the satellite luminosity function, or distribution of mass-to-light ratio at fixed host halo mass. This physically motivated stochasticity has important implications for abundance matching techniques that need to be taken into account and complicates their use. We note that some of this stochasticity is already seen in mildly nonlinear theory (which captures some aspects of the formation history) and manifests itself as a scatter in entropy at large $1+\delta$ in Figure 3.

¹⁸ This assumes star formation in primordial gas. However, other processes such as metal pollution from adjacent galaxies might cause stars to form earlier at lower virial temperatures of $T_{200} < 10^4$ K.

3.2. The Void Phenomenon and the Faint-End Slope of the Galaxy Luminosity and H I-Mass Function

As outlined in Section 1.2, the “void phenomenon” is closely related to the substructure problem. Both show a strong discrepancy in the abundance of dark matter (sub-)halos and paucity of luminous dwarf galaxies that are thought to be hosted by these halos. Before we discuss in detail how blazar heating impacts on dwarf galaxies in voids, we review the latest discussion on the status of the problem itself which has recently been disputed to even exist.

Using a halo occupation distribution model approach, Tinker & Conroy (2009) claim to have explained the problem as they find agreement between luminosity functions, nearest neighbor statistics, and void probability function of faint galaxies. However, a very high resolution simulation of the local volume of 8 Mpc around the Milky Way predicts a factor of 10 more dwarf halos than observed dwarf galaxies in mini-voids with sizes ranging from 1 to 4.5 Mpc, hence reinforcing the “void phenomenon” (Tikhonov & Klypin 2009). While the agreement between theory and observations is good for dwarfs with masses $M_{200m} \gtrsim 10^{10} M_{\odot}$ (corresponding to maximum circular velocities of $v_c \gtrsim 40 \text{ km s}^{-1}$),¹⁹ it fails below, suggesting that Tinker & Conroy (2009) did not sample the relevant mass scales. Moreover, their analysis only demonstrated the self-consistency of the halo occupation distribution model and did not make a comparison between observations and the predictions of Λ CDM.

The discrepancy of the luminosity function on dwarf scales is confirmed by other recent studies that compare the circular velocity function of H I observations with those obtained through dissipationless simulations (Zavala et al. 2009; Zwaan et al. 2010; Trujillo-Gomez et al. 2011). At $v_c < 80 \text{ km s}^{-1}$ (corresponding to $M_{200m} < 10^{11} M_{\odot}$), the latest study finds a slight deviation between theory and observations which amounts to a significant overprediction of more than 10 times the number of observed systems at $v_c = 40 \text{ km s}^{-1}$. While completeness of the observed samples could be an explanation of some of the differences, it is unlikely that it can account for all of the observed effects since the blind H I sample of the HIPASS survey is thought to be complete down to $M_{\text{HI}} < 5.5 \times 10^7 M_{\odot}$ out to a distance of 5 Mpc (Zwaan et al. 2010). Since gas-rich galaxies dominate at the low mass end of the luminosity function, their sample should give an accurate measurement of the abundance of dwarfs if these galaxies contain enough neutral gas to be detected.

While blazar heating significantly increases the entropy, and hence the associated filtering mass, at mean density, it does so more dramatically within the voids. To address how this substantially increased heating manifests itself in the number of void dwarfs, it is instructive to compare the formation timescales of dwarf halos as a function of environment. Void dwarf galaxies are negatively biased and form later than their field or cluster analogs (halos of a given mass tend to be older in clusters and younger in voids for masses smaller than the typical mass scale that is presently entering the nonlinear regime of perturbation growth). This is because for galaxies form-

¹⁹ We quote the values corresponding to the lower bounds in their models that assumed a normalization of the matter power spectrum of $\sigma_8 = 0.9$. This seems to be a conservative choice when comparing to the critical mass threshold of $M_{200m} > 6 \times 10^9 M_{\odot}$ or equivalently $v_c > 35 \text{ km s}^{-1}$ which assumed a $\sigma_8 = 0.75$. For consistency reasons with the literature, in Section 3, this section, we define the virial mass, M_{200m} , as the mass of a sphere enclosing a mean density that is 200 times the *mean* density of the universe.

ing on a large-scale underdense mode, more time has to elapse before these galaxies acquire enough overdensity to decouple from the Hubble expansion and subsequently collapse. Thus, the median redshift of formation for a dwarf galaxy with halo mass $2.4 \times 10^{10} M_{\odot}$ is $z_{\text{form}} = 2.1$ within cluster and 1.6 in voids (Hahn et al. 2007a). Moreover, the distribution of formation times of void galaxies is more sharply confined around the median value than those for clusters, which exhibit a long tail of formation redshifts, extending to $z \sim 6$ (Hahn et al. 2007b). Our estimate of M_C in underdense regions with $1 + \delta = 0.5$ demonstrates that blazar heating prevents the condensation of baryons in halos of masses of $10^{10} M_{\odot}$ at redshifts $z < 2.4$ and $2.8 \times 10^{11} M_{\odot}$ at $z = 0$ (see Figure 7). In particular, halos of $2 \times 10^{10} M_{\odot}$ can be suppressed after $z = 1.8$ which is earlier than the median formation redshift of these galaxies. Depending on the exact definition of voids, this suggests that more than half of the galaxies with masses $M < 2 \times 10^{10} M_{\odot}$, corresponding to a maximum circular velocity of $v_c < 45 \text{ km s}^{-1}$, can be suppressed or severely affected by blazar heating.

Such a preheated entropy floor not only suppresses dwarf formation at late times ($z \lesssim 2$), but may also modify galaxy formation at the low-mass end. Assuming that low-mass halos are embedded in a preheated medium with an entropy floor of 10 keV cm^2 at $z \lesssim 2$ simultaneously matches data at the faint-end slope of the galaxy luminosity function as well as of the H I-mass function (Mo et al. 2005). This heuristic assumption almost exactly coincides with the predictions of our blazar heating models (see Figure 1). As a result of such a preheating, only a fraction of the gas in a proto-galaxy region would be able to cool and be accreted into the final galaxy halo by the present time. If the accreted gas resides in the diffuse phase, it does not lose angular momentum to the dark matter, thereby possibly continuing to form large galaxy discs in low-density environments (Mo & Mao 2002).

In summary, the entropy floor and filtering mass due to blazar heating are dramatically increased in voids as a result of the constant volumetric heating rate, which leads to an inverted temperature-density relation in low-density regions (Paper II). In combination with the later formation epoch of dwarfs at these low densities, this implies a very efficient mechanism for suppressing void dwarf formation in collapsed dark matter halos. Hence, our model provides an elegant physical solution to the void phenomenon described by Peebles (2001).

3.3. Suppression of Dwarfs in Warm Dark Matter Cosmologies

Recent dissipationless Λ CDM simulations produce not only far too many dark matter satellites but also show that the most massive subhalos in simulations of the Milky Way may be too dense to host any of its observed bright satellites with luminosities $L_V > 10^5 L_{\odot}$, i.e., these massive satellites in simulations attain their maximum circular velocity at too small radii in comparison to the observed dwarf satellites (Boylan-Kolchin et al. 2011b,a). These dark subhalos have circular velocities at infall of $30 - 70 \text{ km s}^{-1}$ and infall masses of $(0.2 - 4) \times 10^{10} M_{\odot}$. In principle, this puzzle can be solved (or partially solved) in the following ways: by increasing the stochasticity of galaxy formation on these scales, by reducing the central (dark matter) densities by means of very efficient and violent baryonic feedback processes acting on timescales much faster than the free-fall time, by assuming a total mass of the Milky Way at the lower end of the allowed uncertainty interval (i.e., \sim

$8 \times 10^{11} M_{\odot}$) in combination with a shallower subhalo density profile of the Einasto form as measured in simulations (Vera-Ciro et al. 2012), or by allowing these subhalos to initially form with lower concentrations as would be the case, for example, if the dark matter were made of warm, rather than cold particles (Lovell et al. 2012).

In the limit of heavy WDM particles, e.g., the (sterile) neutrino with $m_{\nu}c^2 \gtrsim 10$ keV that is created in the early universe through mixing with an active neutrino, structure formation proceeds almost indistinguishably from CDM for all current observational probes (Seljak et al. 2006). For smaller masses, however, the free streaming of neutrinos erases all fluctuations on scales smaller than the free streaming length, which is roughly proportional to their temperature and inversely proportional to their mass. Using the Ly α forest power spectrum measured by the Sloan Digital Sky Survey and high-resolution spectroscopy observations in combination with CMB and galaxy clustering constraints still allows for a neutrino with mass $m_{\nu}c^2 > 2.5$ keV (95% c.l.) that decoupled early while in thermal equilibrium (Seljak et al. 2006).²⁰

Dissipationless high-resolution simulations of the evolution of a Galaxy-sized halo have shown that (sub-)halos form and are accreted later onto the main halo in WDM scenarios compared to the standard CDM paradigm, due to the lack of power on small scales in WDM (Macciò & Fontanot 2010). In WDM scenarios with relatively low values of the particle masses, $m_{\nu}c^2 = (2-5)$ keV, there are almost no halos with $z_{\text{form}} \geq 11$, suggesting that the fraction of late-forming dwarfs $z_{\text{form}} \lesssim 1.5$ is increased over the CDM scenario. This late formation epoch of dwarfs reinforces the star formation history problem of Local Group dwarfs: these late-forming dwarfs contain young stellar populations, in direct conflict with the old ages of the stellar populations ($\tau > 10$ Gyr) in all of the Local Group dwarfs (Dolphin et al. 2005; Holtzman et al. 2006; Orban et al. 2008).

However, these are precisely the objects that blazar heating most strongly affects, as the strongly rising filtering mass (or entropy floor) after $z \sim 2$ is able to suppress the population of these late-forming dwarfs. Hence, while in this scenario the free streaming of WDM erases power on scales smaller than dwarfs, blazar heating reconciles the theoretically expected and the observed star formation histories and alleviates standard objections to galaxy formation in WDM cosmologies.

3.4. Impact on the formation of L_* galaxies

In the previous sections, we argued that a blazar-heated entropy floor is able to suppress late dwarf formation in voids and the Milky Way as well as modify the thermodynamical profiles of galaxy groups and clusters. Hence, it is natural to ask whether there would be any effect of blazar heating on L_* galaxy formation which represents the mass scale in between these two extremes. Blazar heating is not powerful enough to raise the mean temperature at $\delta = 0$ of even the most extreme patches above 10^5 K (see Figure 9 in Paper II). The classical criteria of galaxy formation are a short cooling time compared to the dynamical time and to the age of the universe, $t_{\text{cool}} \lesssim H^{-1}$ and $t_{\text{cool}} \lesssim t_{\text{dyn}}$ (Rees & Ostriker 1977; Silk 1977; White & Rees 1978), which are easily fulfilled even in the pres-

²⁰ Since blazar heating dramatically changes the thermal history of the IGM (Paper II) and may be responsible for the inverted temperature-density relation at $z = 2-3$ inferred by high-redshift Ly α studies (Bolton et al. 2008; Viel et al. 2009; Puchwein et al. 2011), it is not clear whether these limits on sterile neutrino properties are weakened in the presence of blazar heating.

ence of blazar heating (see, e.g., Figure 1 in Rees & Ostriker 1977). This implies that these galaxies can radiate away the additional entropy that the gas attained prior to collapse due to blazar heating within a free-fall time. It is, however, interesting to speculate whether the blazar-heated high-entropy gas at low redshift $z < 1$ has any impact on the late accretion of gas into the hot reservoir of baryons from which gas cools and fuels the late time star formation. Higher entropy gas should shock further out in the halo than pre-cooled gas of lower entropy which is denser and can provide a larger ram pressure. Blazar heated gas has high entropy and is more dilute. Hence, it is more easily torqued (by dissipative processes or magnetic fields) and therefore change its angular momentum distribution. Thus, it is implausible that blazar heating dramatically changes the ordinary mode of galaxy formation, but we might anticipate blazar heating to starve the late-time accretion and potentially slow down subsequent star formation. These ideas are subject to verification by numerical simulations of galaxy formation.

4. CONCLUSIONS

TeV blazar heating results in a dramatic increase in the entropy of the IGM following He II reionization around $z \sim 3.5$. Since the IGM entropy evolution is critical for the formation and structure of collapsed objects, blazars heating has a significant impact upon both. We have identified two mass ranges (or classes of objects) for which the TeV blazar-induced entropy floor should have a substantial effect. *Galaxy groups and clusters*, which are forming near the peak entropy injection rate ($z \sim 1$) and exhibit core entropies that are comparable to that implied by blazar heating; and *dwarf galaxies*, which are susceptible to the rapidly rising entropy floor generated by blazars. Below, we describe the consequences for each in more detail.

Galaxy groups and clusters. Immediately after formation, groups at fixed mass should have a continuous distribution of core entropy values, depending on the formation redshift and the temporally variable heating mechanism. The fate of these groups is determined by the ratio of the cooling time, t_{cool} , to the timescale between cluster/group mergers, t_{merger} . If this ratio is smaller than unity, the group can radiate the elevated core entropy away and evolve into a CC which survives the successive hierarchical growth. Alternatively, if $t_{\text{cool}} > t_{\text{merger}}$, merger shocks can gravitationally reprocess the entropy cores and amplify them. Those groups can then evolve into NCC systems. Hence, it is not necessary to produce all of the observed central entropy in the IGM before collapse, but it is also possible to achieve this through gravitational heating, provided there is a certain minimum entropy delivered by some putative heating process corresponding to a minimum t_{cool} . An increasing entropy floor also implies an increasing cooling time, hence the cluster-averaged $t_{\text{cool}}/t_{\text{merger}}$ increases. It follows that systems that evolve into CC systems *today* are on average early-forming, i.e., old systems. In contrast, NCCs are on average young systems.

We argue that systems with a unimodal distribution in core entropy values after group formation should naturally evolve into a bimodal distribution. The reason for this are the two attractor solutions of the group/cluster system, driven by the cooling instability and gravitational reprocessing of (temporally increasing) elevated entropy cores, resulting in the observed populations of CC and NCC systems, respectively. Such an elevated entropy core level in groups might explain a population of X-ray dim groups with low gas fractions. We show

that the core entropy values in CC clusters have a very weak correlation with the mechanical energy and power of X-ray cavities inflated by AGNs. Apparently, PdV work done by the expanding cavities is an inefficient heating process that does not generate much entropy. This strongly suggests that while AGN feedback seems to be critical in stabilizing CC systems, it cannot transform CC into NCC systems (at least on the buoyancy timescale).

Our blazar-induced entropy history seems to be well matched for the formation times of today's groups, but at first sight less so for clusters which form in highly biased regions through mergers of groups. Those had to form even earlier when the entropy for the *average* IGM was still rising with typical values at $z \simeq 1$ of $K_0 \simeq (25-50)$ keV cm 2 . However, two effects positively interfere to counteract the apparently smaller effect of blazar heating in clusters. First, the mass accretion rate is larger for larger systems and at higher redshifts. This suggests that the earlier forming group progenitors of clusters can tolerate a smaller blazar-heated entropy floor after collapse which will then be gravitationally processed at a faster rate and able to counteract the smaller cooling timescales. Second, blazars also turn on first in highly biased regions, and thus the IGM in the vicinity of clusters should experience blazar heating earlier than low-density regions. We speculate that this effect in combination with the faster gravitational reprocessing of an elevated entropy core in denser regions would help in propagating the effect of blazar heating to the scale of massive clusters.

Changing the thermodynamic structure of groups and potentially clusters has also an impact on the SZ power spectrum which is sensitive to cosmological parameters such as σ_8 and the thermal pressure profile. An increased core entropy level implies a smoother pressure profile and hence decreases the power in the SZ power spectrum on scales smaller than 3' which are probed by current experiments such as SPT and ACT. Lowering the astrophysical signal can allow for larger values of σ_8 , potentially reducing some (minor) tension with the data especially after allowing for a contribution due to patchy reionization.

Dwarf galaxy formation. We demonstrate that the redshift-dependent entropy floor increases the characteristic halo mass, M_C , below which dwarf galaxies cannot form by a factor of 6.5–15 for the cosmic mean ($\delta = 0$) and by a factor of 25–70 for voids ($\delta = -0.5$). The range indicates the uncertainty in the number of blazars that contribute to the heating and the upper envelope matches the observations of an inverted temperature-density relation in the Ly α forest. The increase of M_C prevents the formation of late-forming dwarf galaxies ($z \lesssim 2$) with masses ranging from 10^{10} to 10^{11} M $_{\odot}$ for redshifts $z \sim 2$ to 0, respectively. This may resolve the “missing satellites problem” in the Milky Way, i.e., the low observed abundance of dwarf satellites compared to CDM simulations. It also brings the observed early star formation histories of Local Group dwarfs into agreement with galaxy formation models that predicted a population of late-forming objects, in conflict with the data. At the same time, it provides a plausible explanation for the “void phenomenon” which is the apparent discrepancy of the number of dwarfs in low-density regions in CDM simulations and the paucity of those in observations. Blazar heating suppresses the formation of galaxies within existing dwarf halos of masses $< 3 \times 10^{10}$ M $_{\odot}$ with a maximum circular velocity < 60 km s $^{-1}$ for $z \lesssim 2$. Additionally, the phenomenology of such a preheating mechanism matches heuristic assumptions that were adopted to match the faint-end slope of the galaxy

luminosity function as well as of the H I-mass function, in particular for low-density environments.

We conclude that the presented scenario of blazar heating holds the promise for solving some of the most outstanding problems in high-energy gamma-ray astrophysics, the IGM as probed by the high-redshift Ly α forest, the formation of galaxies, and clusters of galaxies. At the same time, it provides an astrophysical solution to these problems such as the “missing satellite problem” or the “void phenomenon” which have been claimed to require new physics beyond the concordance cosmological Λ CDM model.

We thank Tom Abel, Marco Ajello, Marcelo Alvarez, Arif Babul, Roger Blandford, James Bolton, Mike Boylan-Kolchin, Luigi Costamante, Andrei Gruzinov, Peter Goldreich, Martin Haehnelt, Andrey Kravtsov, Hojun Mo, Ue-li Pen, Ewald Puchwein, Volker Springel, Chris Thompson, Matteo Viel, Marc Voit, and Risa Wechsler for useful discussions. We are indebted to Peng Oh for his encouragement and useful suggestions. We thank Steve Furlanetto for kindly providing technical expertise. We also thank the referee for a thorough reading of the manuscript and for his constructive comments. These computations were performed on the Sunnyvale cluster at CITA. A.E.B. and P.C. are supported by CITA. A.E.B. gratefully acknowledges the support of the Beatrice D. Tremaine Fellowship. C.P. gratefully acknowledges financial support of the Klaus Tschira Foundation and would furthermore like to thank KITP for their hospitality during the galaxy cluster workshop. This research was supported in part by the National Science Foundation under Grant No. NSF PHY05-51164.

APPENDIX

A. COMPUTING THE FILTERING MASS GENERALLY

Here we collect the relevant details for computing the filtering mass, M_F , associated with a particular cosmological and thermodynamic evolution of the IGM. Recall that the filtering wavenumber is related to the Jeans wavenumber via

$$\frac{1}{k_F^2(t)} = \frac{1}{D_+(t)} \int_0^t dt' a^2(t') \frac{\ddot{D}_+(t') + 2H(t')\dot{D}_+(t')}{k_J^2(t')} \times \int_{t'}^t \frac{dt''}{a^2(t'')} . \quad (13)$$

We can simplify Equation (13) by noting that the linear growth function obeys the following equation

$$\ddot{D}_+(t) + 2H(t)\dot{D}_+(t) = 4\pi G \bar{\rho} D_+(t), \quad (A1)$$

in which the Hubble function is given by

$$\frac{H^2(a)}{H_0^2} = E^2(a) = a^{-3}\Omega_m + a^{-2}(1 - \Omega_m - \Omega_{\Lambda}) + \Omega_{\Lambda} . \quad (A2)$$

Substituting the integration variable in Equation (13) by the scale factor a , we obtain

$$\frac{1}{k_F^2(a)} = \frac{1}{D_+(a)} \int_0^a da' \frac{c_s^2(a')}{H_0^2} \frac{D_+(a')}{a'E(a')} \int_{a'}^a \frac{da''}{a''^3 E(a'')} . \quad (A3)$$

Following Carroll et al. (1992), the linear growth function can be computed by

$$D_+(a) = \frac{5}{2} \Omega_m E(a) \int_0^a da' \frac{1}{a'^3 E^3(a')} . \quad (A4)$$

For any given temperature evolution of the IGM (which enters via the sound speed), we can compute a redshift evolution of the filtering scale from Equation (A3). It is convenient to define a filtering mass by analogy with the Jeans mass:

$$M_F(a) \equiv \frac{4\pi}{3} \bar{\rho}(a) \left(\frac{2\pi a}{k_F(a)} \right)^3. \quad (\text{A5})$$

Finally, using a definition for the entropy of Equation (1), we can derive the filtering scale $\lambda_F = 2\pi a/k_F$ in terms of the entropy,

$$\frac{1}{k_F^2(a)} = \frac{A_0}{D_+(a)} \int_0^a da' K(a') \frac{D_+(a')}{a'^3 E(a')} \int_{a'}^a \frac{da''}{a''^3 E(a'')}, \quad (\text{A6})$$

$$A_0 = \frac{5}{3} \left(\frac{3\Omega_m}{8\pi G H_0} \right)^{2/3}. \quad (\text{A7})$$

B. FILTERING MASS IN AN EINSTEIN-DE SITTER UNIVERSE

For sufficiently early times, all matter-dominated Friedmann-Lemaître model universes asymptotically approach an Einstein-de Sitter universe ($z \gtrsim 2$ for our Λ CDM universe), for which the cosmological constant $\Lambda = 0$ and the curvature $\Omega_k = 0$. For such a universe, the growth function $D_+(a) \propto a$ which allows us to considerably simplify Equation (A3), yielding

$$\frac{1}{k_F^2(a)} = \frac{3}{a} \int_{a_{\min}}^a da' \frac{1}{k_F^2(a')} \left[1 - \left(\frac{a'}{a} \right)^{1/2} \right]. \quad (\text{B1})$$

Note that we replaced the unphysical lower integration limit 0 by the recombination scale factor, a_{\min} , as baryon perturbations can only start to grow after recombination. Employing the relation of Equation (A5) we can rewrite this to give

$$M_F^{2/3}(a) = \frac{3}{a} \int_{a_{\min}}^a da' M_J^{2/3}(a') \left[1 - \left(\frac{a'}{a} \right)^{1/2} \right]. \quad (\text{B2})$$

After recombination, the residual electron density couples the gas temperature still to that of the CMB via Compton interactions. Hence, we expect the gas temperature to scale as $T \propto a^{-1}$. At $z \simeq 150$, the Compton interaction rate drops below the Hubble expansion rate such that the gas experiences adiabatic expansion with $T \propto a^{-2}$. Hence for the time after matter-radiation equality, where $\bar{\rho} = \bar{\rho}_0 a^{-3}$, we can write in general

$$\frac{T(a)}{T_0} = \left(\frac{a}{a_0} \right)^{-\alpha}, \quad \text{and} \quad \frac{k_J(a)}{k_{J,0}} = \left(\frac{a}{a_0} \right)^{(\alpha-1)/2}. \quad (\text{B3})$$

This definition for the Jeans scale implies a Jeans mass at a fiducial scale factor, a_0 , of

$$M_{J,0} \equiv \frac{4\pi}{3} \bar{\rho}(a_0) \left(\frac{2\pi a_0}{k_{J,0}} \right)^3. \quad (\text{B4})$$

If we substitute Equation (B3) into Equation (B2), we can obtain an analytical solution for the filtering mass before reionization,

$$M_F(a) = M_{J,0} \left\{ \frac{3}{a} \int_{a_{\min}}^a da' \left(\frac{a}{a_0} \right)^{1-\alpha} \left[1 - \left(\frac{a'}{a} \right)^{1/2} \right] \right\}^{3/2}. \quad (\text{B5})$$

Evaluating this integral for the two regimes before and after the freezeout of Compton interactions of the gas with the CMB photons, yields the following analytic solutions,

$$M_F(a) = M_{J,0} \begin{cases} \left[\frac{6a_0}{a} \left(\sqrt{\frac{a_{\min}}{a}} - 1 + \frac{1}{2} \ln \frac{a}{a_{\min}} \right) \right]^{3/2} & \text{for } 150 < z < 1100 \\ \left[1 - 3 \frac{a_{\min}}{a} + 2 \left(\frac{a_{\min}}{a} \right)^{3/2} \right]^{3/2} & \text{for } z_{\text{reion}} \leq z \leq 150. \end{cases} \quad (\text{B6})$$

REFERENCES

Abdo, A. A., et al. 2010, *Phys. Rev. Lett.*, 104, 101101
 Babul, A., Balogh, M. L., Lewis, G. F., & Poole, G. B. 2002, *MNRAS*, 330, 329
 Balogh, M. L., Babul, A., & Patton, D. R. 1999, *MNRAS*, 307, 463
 Barkana, R., & Loeb, A. 1999, *ApJ*, 523, 54
 Battaglia, N., Bond, J. R., Pfrommer, C., & Sievers, J. L. 2011, arXiv:1109.3711
 Battaglia, N., Bond, J. R., Pfrommer, C., Sievers, J. L., & Sijacki, D. 2010, *ApJ*, 725, 91
 Benson, A. J., & Madau, P. 2003, *MNRAS*, 344, 835
 Bialek, J. J., Evrard, A. E., & Mohr, J. J. 2001, *ApJ*, 555, 597
 Bîrzan, L., Rafferty, D. A., McNamara, B. R., Wise, M. W., & Nulsen, P. E. J. 2004, *ApJ*, 607, 800
 Bolton, J. S., Viel, M., Kim, T., Haehnelt, M. G., & Carswell, R. F. 2008, *MNRAS*, 386, 1131
 Booth, C. M., & Schaye, J. 2009, *MNRAS*, 398, 53
 Borgani, S., Finoguenov, A., Kay, S. T., Ponman, T. J., Springel, V., Tozzi, P., & Voit, G. M. 2005, *MNRAS*, 361, 233
 Borgani, S., Governato, F., Wadsley, J., Menci, N., Tozzi, P., Lake, G., Quinn, T., & Stadel, J. 2001, *ApJ*, 559, L71
 Borgani, S., & Kravtsov, A. 2009, arXiv:0906.4370
 Borgani, S., & Viel, M. 2009, *MNRAS*, 392, L26
 Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011a, arXiv:1111.2048
 —, 2011b, *MNRAS*, 415, L40
 Bret, A. 2009, *ApJ*, 699, 990
 Bret, A., Firpo, M., & Deutsch, C. 2004, *Phys. Rev. E*, 70, 046401
 —, 2005, *Physical Review Letters*, 94, 115002
 Bret, A., Gremillet, L., & Dieckmann, M. E. 2010, *Physics of Plasmas*, 17, 120501
 Broderick, A. E., Chang, P., & Pfrommer, C. 2012, *ApJ* in print, arXiv:1106.5494
 Bryan, G. L., & Norman, M. L. 1998, *ApJ*, 495, 80
 Bryan, G. L., & Voit, G. M. 2001, *ApJ*, 556, 590
 Bullock, J. S., Kravtsov, A. V., & Weinberg, D. H. 2000, *ApJ*, 539, 517
 —, 2001, *ApJ*, 548, 33
 Busha, M. T., Alvarez, M. A., Wechsler, R. H., Abel, T., & Strigari, L. E. 2010, *ApJ*, 710, 408
 Carroll, S. M., Press, W. H., & Turner, E. L. 1992, *ARA&A*, 30, 499
 Cavadini, M., Salvaterra, R., & Haardt, F. 2011, arXiv:1105.4613
 Cavagnolo, K. W., Donahue, M., Voit, G. M., & Sun, M. 2008, *ApJ*, 683, L107
 —, 2009, *ApJS*, 182, 12
 Chang, P., Broderick, A. E., & Pfrommer, C. 2012, *ApJ* in print, arXiv:1106.5504
 Chiu, W. A., Gnedin, N. Y., & Ostriker, J. P. 2001, *ApJ*, 563, 21
 Churazov, E., Brüggen, M., Kaiser, C. R., Böhringer, H., & Forman, W. 2001, *ApJ*, 554, 261
 Croft, R. A. C., Di Matteo, T., Davé, R., Hernquist, L., Katz, N., Fardal, M. A., & Weinberg, D. H. 2001, *ApJ*, 557, 67
 Dai, X., Bregman, J. N., Kochanek, C. S., & Rasia, E. 2010, *ApJ*, 719, 119
 Dalal, N., & Kochanek, C. S. 2002, *ApJ*, 572, 25
 Dalanton, J. J., & Hogan, C. J. 2001, *ApJ*, 561, 35
 Dekel, A., & Woo, J. 2003, *MNRAS*, 344, 1131
 Dermer, C. D., Cavadini, M., Razzaque, S., Finke, J. D., Chiang, J., & Lott, B. 2011, *ApJ*, 733, L21
 Dijkstra, M., Haiman, Z., Rees, M. J., & Weinberg, D. H. 2004, *ApJ*, 601, 666
 Dolag, K., Kachelriess, M., Ostapchenko, S., & Tomàs, R. 2011, *ApJ*, 727, L4
 Dolphин, A. E., Weisz, D. R., Skillman, E. D., & Holtzman, J. A. 2005, astro-ph/050643

Donahue, M., de Messières, G. E., O'Connell, R. W., Voit, G. M., Hoffer, A., McNamara, B. R., & Nulsen, P. E. J. 2011, *ApJ*, 732, 40

Dubois, Y., Devriendt, J., Slyz, A., & Teyssier, R. 2010, *MNRAS*, 409, 985

Dunkley, J., et al. 2011, *ApJ*, 739, 52

Dursi, L. J., & Pfrommer, C. 2008, *ApJ*, 677, 993

Efstathiou, G. 1992, *MNRAS*, 256, 43P

Efstathiou, G., & Migliaccio, M. 2011, arXiv:1106.3208

Eisenstein, D. J., & Hu, W. 1998, *ApJ*, 496, 605

Enßlin, T., Pfrommer, C., Miniati, F., & Subramanian, K. 2011, *A&A*, 527, 99

Enßlin, T. A., & Vogt, C. 2006, *A&A*, 453, 447

Evrard, A. E., & Henry, J. P. 1991, *ApJ*, 383, 95

Evrard, A. E., Metzler, C. A., & Navarro, J. F. 1996, *ApJ*, 469, 494

Fang, W., & Haiman, Z. 2008, *ApJ*, 680, 200

Farrar, G. R., & Peebles, P. J. E. 2004, *ApJ*, 604, 1

Fenner, Y., Gibson, B. K., Gallino, R., & Lugaro, M. 2006, *ApJ*, 646, 184

Font, A. S., et al. 2011, arXiv:1103.0024

Fowler, J. W., et al. 2010, *ApJ*, 722, 1148

Frenk, C. S., et al. 1999, *ApJ*, 525, 554

Gnedin, N. Y. 2000, *ApJ*, 542, 535

Gnedin, N. Y., & Hui, L. 1998, *MNRAS*, 296, 44

Gnedin, N. Y., Tassis, K., & Kravtsov, A. V. 2009, *ApJ*, 697, 55

Gottlöber, S., Klypin, A., & Kravtsov, A. V. 2001, *ApJ*, 546, 223

Guo, F., & Oh, S. P. 2008, *MNRAS*, 384, 251
—. 2009, *MNRAS*, 400, 1992

Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, *MNRAS*, 404, 1111

Hahn, O., Carollo, C. M., Porciani, C., & Dekel, A. 2007a, *MNRAS*, 381, 41

Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007b, *MNRAS*, 375, 489

Haiman, Z., Abel, T., & Rees, M. J. 2000, *ApJ*, 534, 11

Haiman, Z., Rees, M. J., & Loeb, A. 1997, *ApJ*, 476, 458

Hicks, A. K., et al. 2008, *ApJ*, 680, 1022

Hinton, J. A., & Hofmann, W. 2009, *ARA&A*, 47, 523

Hoeft, M., Yepes, G., Gottlöber, S., & Springel, V. 2006, *MNRAS*, 371, 401

Holtzman, J. A., Afonso, C., & Dolphin, A. 2006, *ApJS*, 166, 534

Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, *ApJ*, 654, 731

Hui, L., & Gnedin, N. Y. 1997, *MNRAS*, 292, 27

Hui, L., & Haiman, Z. 2003, *ApJ*, 596, 9

Iliev, I. T., Mellema, G., Pen, U., Bond, J. R., & Shapiro, P. R. 2008, *MNRAS*, 384, 863

Iliev, I. T., Pen, U., Bond, J. R., Mellema, G., & Shapiro, P. R. 2007, *ApJ*, 660, 933

Inoue, Y., & Totani, T. 2009, *ApJ*, 702, 523

Jubelgas, M., Springel, V., Enßlin, T., & Pfrommer, C. 2008, *A&A*, 481, 33

Kaiser, N. 1986, *MNRAS*, 222, 323
—. 1991, *ApJ*, 383, 104

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, *MNRAS*, 264, 201

Kaufmann, T., Wheeler, C., & Bullock, J. S. 2007, *MNRAS*, 382, 1187

Keisler, R., et al. 2011, *ApJ*, 743, 28

Kitayama, T., & Ikeuchi, S. 2000, *ApJ*, 529, 615

Kneiske, T. M., & Mannheim, K. 2008, *A&A*, 479, 41

Komatsu, E., et al. 2011, *ApJS*, 192, 18

Koposov, S. E., Yoo, J., Rix, H.-W., Weinberg, D. H., Macciò, A. V., & Escudé, J. M. 2009, *ApJ*, 696, 2179

Kravtsov, A. 2010, *Advances in Astronomy*, 2010

Kravtsov, A. V., Gnedin, O. Y., & Klypin, A. A. 2004, *ApJ*, 609, 482

Kulsrud, R., & Pearce, W. P. 1969, *ApJ*, 156, 445

Kunz, M. W., Schekochihin, A. A., Cowley, S. C., Binney, J. J., & Sanders, J. S. 2011, *MNRAS*, 410, 2446

Lemoine, M., & Pelletier, G. 2010, *MNRAS*, 402, 321

Lidz, A., Faucher-Giguère, C., Dall'Aglio, A., McQuinn, M., Fechner, C., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2010, *ApJ*, 718, 199

Lovell, M. R., et al. 2012, *MNRAS*, 420, 2318

Lueker, M., et al. 2010, *ApJ*, 719, 1045

Lyutikov, M. 2006, *MNRAS*, 373, 73

Mac Low, M.-M., & Ferrara, A. 1999, *ApJ*, 513, 142

Macciò, A. V., & Fontanot, F. 2010, *MNRAS*, 404, L16

Macciò, A. V., Kang, X., Fontanot, F., Somerville, R. S., Koposov, S., & Monaco, P. 2010, *MNRAS*, 402, 1995

Mahdavi, A., Böhringer, H., Geller, M. J., & Ramella, M. 2000, *ApJ*, 534, 114

Mao, S., Jing, Y., Ostriker, J. P., & Weller, J. 2004, *ApJ*, 604, L5

Markevitch, M. 1998, *ApJ*, 504, 27

Marriage, T. A., et al. 2011, *ApJ*, 737, 61

Mashchenko, S., Wadsley, J., & Couchman, H. M. P. 2008, *Science*, 319, 174

McCarthy, I. G., Babul, A., Bower, R. G., & Balogh, M. L. 2008, *MNRAS*, 386, 1309

McCarthy, I. G., et al. 2010, *MNRAS*, 406, 822

McCourt, M., Sharma, P., Quataert, E., & Parrish, I. J. 2012, *MNRAS*, 419, 3319

McNamara, B. R., & Nulsen, P. E. J. 2007, *ARA&A*, 45, 117

Miniati, F., Ryu, D., Kang, H., Jones, T. W., Cen, R., & Ostriker, J. P. 2000, *ApJ*, 542, 608

Mo, H. J., & Mao, S. 2002, *MNRAS*, 333, 768

Mo, H. J., Yang, X., van den Bosch, F. C., & Katz, N. 2005, *MNRAS*, 363, 1155

Muñoz, J. A., Madau, P., Loeb, A., & Diemand, J. 2009, *MNRAS*, 400, 1593

Narumoto, T., & Totani, T. 2006, *ApJ*, 643, 81

Neronov, A., & Vovk, I. 2010, *Science*, 328, 73

Nickerson, S., Stinson, G., Couchman, H. M. P., Bailin, J., & Wadsley, J. 2011, *MNRAS*, 415, 257

Nusser, A., Gubser, S. S., & Peebles, P. J. 2005, *Phys. Rev. D*, 71, 083505

Oh, S. P., & Benson, A. J. 2003, *MNRAS*, 342, 664

Okamoto, T., Gao, L., & Theuns, T. 2008, *MNRAS*, 390, 920

Orban, C., Gnedin, O. Y., Weisz, D. R., Skillman, E. D., Dolphin, A. E., & Holtzman, J. A. 2008, *ApJ*, 686, 1030

Parrish, I. J., Quataert, E., & Sharma, P. 2010, *ApJ*, 712, L194

Paul, S., Iapichino, L., Miniati, F., Bagchi, J., & Mannheim, K. 2011, *ApJ*, 726, 17

Peebles, P. J. E. 2001, *ApJ*, 557, 495

Pen, U. 1999, *ApJ*, 510, L1

Pfrommer, C., & Dursi, L. J. 2010, *Nature Physics*, 6, 520

Pfrommer, C., Springel, V., Enßlin, T. A., & Jubelgas, M. 2006, *MNRAS*, 367, 1113

Ponman, T. J., Cannon, D. B., & Navarro, J. F. 1999, *Nature*, 397, 135

Poole, G. B., Babul, A., McCarthy, I. G., Sanderson, A. J. R., & Fardal, M. A. 2008, *MNRAS*, 391, 1163

Pratt, G. W., et al. 2010, *A&A*, 511, 85

Puchwein, E., Pfrommer, C., Springel, V., Broderick, A. E., & Chang, P. 2011, arXiv:1107.3837

Puchwein, E., Sijacki, D., & Springel, V. 2008, *ApJ*, 687, L53

Quinn, T., Katz, N., & Efstathiou, G. 1996, *MNRAS*, 278, L49

Rafferty, D. A., McNamara, B. R., Nulsen, P. E. J., & Wise, M. W. 2006, *ApJ*, 652, 216

Rees, M. J., & Ostriker, J. P. 1977, *MNRAS*, 179, 541

Robertson, B. E., & Kravtsov, A. V. 2008, *ApJ*, 680, 1083

Ruszkowski, M., & Oh, S. P. 2010, *ApJ*, 713, 1332

Ryu, D., Kang, H., Hallman, E., & Jones, T. W. 2003, *ApJ*, 593, 599

Sanderson, A. J. R., O'Sullivan, E., & Ponman, T. J. 2009, *MNRAS*, 395, 764

Scannapieco, E., Thacker, R. J., & Davis, M. 2001, *ApJ*, 557, 605

Seljak, U., Makarov, A., McDonald, P., & Trac, H. 2006, *Physical Review Letters*, 97, 191303

Shapiro, P. R., Iliev, I. T., & Raga, A. C. 2004, *MNRAS*, 348, 753

Shirokoff, E., et al. 2011, *ApJ*, 736, 61

Sigward, F., Ferrara, A., & Scannapieco, E. 2005, *MNRAS*, 358, 755

Sijacki, D., Pfrommer, C., Springel, V., & Enßlin, T. A. 2008, *MNRAS*, 387, 1403

Sijacki, D., & Springel, V. 2006, *MNRAS*, 366, 397

Sijacki, D., Springel, V., Di Matteo, T., & Hernquist, L. 2007, *MNRAS*, 380, 877

Silk, J. 1977, *ApJ*, 211, 638

Somerville, R. S. 2002, *ApJ*, 572, L23

Spergel, D. N., & Steinhardt, P. J. 2000, *Physical Review Letters*, 84, 3760

Stanek, R., Rasia, E., Evrard, A. E., Pearce, F., & Gazzola, L. 2010, *ApJ*, 715, 1508

Sun, M., Voit, G. M., Donahue, M., Jones, C., Forman, W., & Vikhlinin, A. 2009, *ApJ*, 693, 1142

Sunyaev, R. A., & Zel'dovich, I. B. 1972, *Comments Astrophys. Space Phys.*, 4, 173

Sunyaev, R. A., & Zeldovich, I. B. 1980, *ARA&A*, 18, 537

Sutherland, R. S., & Dopita, M. A. 1993, *ApJS*, 88, 253

Takahashi, K., Mori, M., Ichiki, K., & Inoue, S. 2012, *ApJ*, 744, L7

Tassis, K., Kravtsov, A. V., & Gnedin, O. Y. 2008, *ApJ*, 672, 888

Tavecchio, F., Ghisellini, G., Bonnoli, G., & Foschini, L. 2011, *MNRAS*, 414, 3566

Tavecchio, F., Ghisellini, G., Foschini, L., Bonnoli, G., Ghirlanda, G., & Coppi, P. 2010, *MNRAS*, 406, L70

Taylor, A. M., Vovk, I., & Neronov, A. 2011, *A&A*, 529, A144

Teyssier, R., Moore, B., Martizzi, D., Dubois, Y., & Mayer, L. 2011, *MNRAS*, 618

Thoul, A. A., & Weinberg, D. H. 1996, *ApJ*, 465, 608

Tikhonov, A. V., & Klypin, A. 2009, *MNRAS*, 395, 1915

Tinker, J. L., & Conroy, C. 2009, *ApJ*, 691, 633

Tozzi, P., & Norman, C. 2001, *ApJ*, 546, 63

Trac, H., Bode, P., & Ostriker, J. P. 2011, *ApJ*, 727, 94

Trujillo-Gomez, S., Klypin, A., Primack, J., & Romanowsky, A. J. 2011, *ApJ*, 742, 16

Uhlig, M., Pfrommer, C., M., S., Nath, B., Enßlin, T. A., & Springel, V. 2012, *MNRAS*, subm.

Vanderlinde, K., et al. 2010, *ApJ*, 722, 1180

Venters, T. M. 2010, *ApJ*, 710, 1530

Vera-Ciro, C. A., Helmi, A., Starkenburg, E., & Breddels, M. A. 2012, arXiv:1202.6061

Viel, M., Bolton, J. S., & Haehnelt, M. G. 2009, *MNRAS*, 399, L39

Vikhlinin, A., Burenin, R., Forman, W. R., Jones, C., Hornstrup, A., Murray, S. S., & Quintana, H. 2007, in *Heating versus Cooling in Galaxies and Clusters of Galaxies*, ed. H. Böhringer, G. W. Pratt, A. Finoguenov, & P. Schuecker , 48

Vikhlinin, A., Kravtsov, A., Forman, W., Jones, C., Markevitch, M., Murray, S. S., & Van Speybroeck, L. 2006, *ApJ*, 640, 691

Voit, G. M. 2005, *Reviews of Modern Physics*, 77, 207

Voit, G. M., Balogh, M. L., Bower, R. G., Lacey, C. G., & Bryan, G. L. 2003, *ApJ*, 593, 272

Voit, G. M., & Bryan, G. L. 2001a, *ApJ*, 551, L139

—. 2001b, *Nature*, 414, 425

Voit, G. M., Bryan, G. L., Balogh, M. L., & Bower, R. G. 2002, *ApJ*, 576, 601

Voit, G. M., Kay, S. T., & Bryan, G. L. 2005, *MNRAS*, 364, 909

Vovk, I., Taylor, A. M., Semikoz, D., & Neronov, A. 2012, *ApJ*, 747, L14

Wadepuhl, M., & Springel, V. 2011, *MNRAS*, 410, 1975

Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, *ApJ*, 568, 52

White, S. D. M., & Rees, M. J. 1978, *MNRAS*, 183, 341

Zavala, J., Jing, Y. P., Faltenbacher, A., Yepes, G., Hoffman, Y., Gottlöber, S., & Catinella, B. 2009, *ApJ*, 700, 1779

Zentner, A. R., & Bullock, J. S. 2003, *ApJ*, 598, 49

Zhao, D. H., Jing, Y. P., Mo, H. J., & Börner, G. 2009, *ApJ*, 707, 354

Zwaan, M. A., Meyer, M. J., & Staveley-Smith, L. 2010, *MNRAS*, 403, 1969